(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
18 December 2003 (18.12.2003)

PCT

(10) International Publication Number
WO 03/104427 A2

(51) International Patent Classification[7]: C12N

(21) International Application Number: PCT/US03/18239

(22) International Filing Date: 5 June 2003 (05.06.2003)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/386,653 5 June 2002 (05.06.2002) US

(71) Applicant *(for all designated States except US)*: **CASE WESTERN RESERVE UNIVERSITY** [US/US]; 10900 Euclid Avenue, Cleveland, OH 44106-7219 (US).

(72) Inventor; and

(75) Inventor/Applicant *(for US only)*: **MARKOWITZ, Sanford, D.** [US/US]; 28950 Bolingbrook Road, Pepper Pike, OH 44124 (US).

(74) Agents: **VARMA, Anita** et al.; Ropes & Gray LLP, One International Place, Boston, MA 02110-2624 (US).

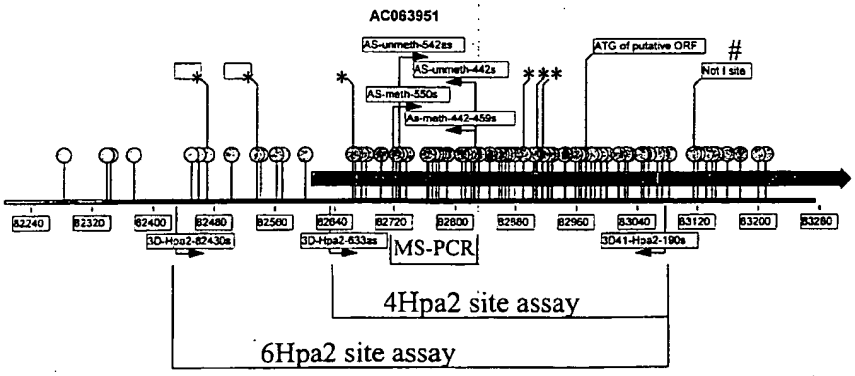
(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO, SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: METHODS AND COMPOSITIONS FOR DETECTING CANCERS



* Denotes a Hpa2 site

\# Denotes the Not1 site

(57) Abstract: In certain aspects, the invention provides isolated SLC5A8 (also referred to as Huil) nucleic acid molecules, which encode novel sodium solute symporter members. The invention also provides antisense nucleic acid molecules, recombinant expression vectors containing SLC5A8 nucleic acid molecules, host cells into which the expression vectors have been introduced, and nonhuman transgenic animals in which a SLC5A8 gene has been introduced or disrupted. The invention still further provides isolated SLC5A8 proteins, fusion proteins, antigenic peptides, and anti-SLC5A8 antibodies. Diagnostic methods utilizing compositions of the invention are also provided. In other aspects, the invention provides methods and compositions for detecting and treating SLC5A8-associated cancer. Differential methylation of the SLC5A8 nucleotide sequences has been observed in SLC5A8-associated cancer, such as colon cancer, breast cancer, thyroid cancer, or stomach cancer.

WO 03/104427 A2

METHODS AND COMPOSITIONS FOR DETECTING CANCERS

Cross-Reference to Related Applications

This application claims the benefit of priority of U.S. Provisional Application No. 60/386,653 filed June 5, 2002, the specification of which is incorporated by reference herein in its entirety.

Funding

Work described herein was supported by National Institutes of Health Grant R01CA 67409. The United States Government has certain rights in the invention.

Background

In 2001, over 1.2 million new cases of human cancer will be diagnosed and over 0.5 million people will die from cancer (American Cancer Society estimate). Despite this, more people than ever are living with and surviving cancer. In 1997, for example, approximately 8.9 million living Americans had a history of cancer (National Cancer Institute estimate). People are more likely to survive cancer if the disease is diagnosed at an early stage of development, since treatment at that time is more likely to be successful. Early detection depends upon availability of high-quality methods. Such methods are also useful for determining patient prognosis, selecting therapy, monitoring response to therapy and selecting patients for additional therapy. Consequently, there is a need for cancer diagnostic methods that are specific, accurate, minimally invasive, technically simple and inexpensive.

Colorectal cancer (cancer of the colon or rectum) is one particularly important type of human cancer. Colorectal cancer is the second most common cause of cancer mortality in adult Americans (Landis, et al., 1999, CA Cancer J Clin, 49:8-31). Approximately 40% of individuals with colorectal cancer die. In 2001, it is estimated that there will be 135,400 new cases of colorectal cancer (98,200 cases of colon and 37,200 cases of rectal cancer) and 56,700 deaths (48,000 colon cancer and 8,800 rectal cancer deaths) from the disease (American Cancer Society). As with other cancers, these rates can be decreased by improved methods for diagnosis. Although methods for detecting colon cancer exist, the methods are not ideal. Digital rectal exams (i.e., manual probing of rectum by a physician), for example, although

relatively inexpensive, are unpleasant and can be inaccurate. Fecal occult blood testing (i.e., detection of blood in stool) is nonspecific because blood in the stool has multiple causes. Colonoscopy and sigmoidoscopy (i.e., direct examination of the colon with a flexible viewing instrument) are both uncomfortable for the patient and expensive. Double-contrast barium enema (i.e., taking X-rays of barium-filled colon) is also an expensive procedure, usually performed by a radiologist.

Other cancers such as breast cancer, thyroid cancer and stomach cancer, cause significant public health problem as well. For example, thyroid cancer is the most common endocrine malignancy. In the United States, there are approximately 14,000 new patients and 1,100 deaths per year (Shah et al., 1995, CA Cancer J Clin 45:352-68). Because of the disadvantages of existing methods for detecting and treating cancer, new methods and tools in cancer diagnosis and cancer therapy are needed.

Summary of the Invention

In accordance with the present invention, new diagnostic tools and methods for detecting cancer (e.g., colon cancer, breast cancer, thyroid cancer, or stomach cancer) are provided. In certain aspects, the invention is based in part on the discovery of a novel polynucleotide sequence encoding a novel sodium/solute symporter-like protein (SLC5A8). Applicants previously referred to the SLC5A8 gene as the "Hui1" gene.

In one embodiment, the invention provides an isolated polypeptide comprising an amino acid sequence selected from the group consisting of: a) an amino acid sequence at least 95% identical to SEQ ID NO: 1; and b) an amino acid sequence encoded by a nucleic acid that hybridizes under high stringency conditions to a nucleic acid of any one of SEQ ID NOs: 3 or 4, wherein said polypeptide is a cell surface protein. The subject polypeptide comprises a transmembrane domain as set forth in any one of SEQ ID NOs: 19-31. The present invention contemplates the subject polypeptide as a sodium symporter.

In another embodiment, the invention provides an isolated antibody or fragment thereof, which is specifically immunoreactive with an epitope of a SCL5A8 protein sequence as set forth in SEQ ID NO: 1. The antibody of the invention can be selected from the group consisting of: a polyclonal antibody, a monoclonal antibody, an Fab fragment and a single chain antibody. Optionally, the antibody is labeled with a detectable label.

In another embodiment, the invention provides an isolated SCL5A8 nucleic acid selected from the group consisting of: a) a nucleic acid comprising the nucleotide sequence of SEQ ID NO: 2, or a complement thereof; b) a nucleic acid molecule that encodes a polypeptide comprising the amino acid sequence at least 95% identical to the amino acid sequence of SEQ ID NO: 7; and c) a nucleic acid molecule that hybridizes under stringent conditions to SEQ ID NO: 2. Optionally, the nucleic acid of the invention further comprises a vector nucleic acid sequence. In certain embodiments, the invention provides a kit comprising the SLC5A8 nucleic acid probes or primers and instructions for use.

In another embodiment, the invention provides a host cell which contains the subject SCL5A8 nucleic acid of the invention. In another embodiment, the invention provides a method for producing the subject polypeptide, comprising culturing the host cell under conditions in which the subject nucleic acid molecule is expressed.

In another embodiment, the invention provides a method for detecting the presence of the subject SCL5A8 polypeptide in a sample, comprising: a) contacting the sample with an antibody which selectively binds to the polypeptide of claim 1; and b) determining whether the antibody binds to the polypeptide in the sample.

In another embodiment, the invention provides a kit for detecting a human SCL5A8 polypeptide comprising: (i) an antibody of claim 2; and (ii) a detectable label for detecting said antibody.

In another embodiment, the invention provides a method for detecting the presence of the SCL5A8 nucleic acid in a sample, comprising: a) contacting the sample with an SCL5A8 probe or primer; and b) determining whether the probe or primer binds to a SCL5A8 nucleic acid in the sample.

In another embodiment, the invention provides a method for identifying a compound which binds to the SCL5A8 polypeptide, comprising: a) contacting the polypeptide, or a cell expressing the SCL5A8 polypeptide, with a test compound; andb) determining whether the polypeptide binds to the test compound.

In another embodiment, the invention provides a method for modulating the activity of the SCL5A8 polypeptide, comprising contacting the polypeptide or a cell expressing the

polypeptide with a compound which binds to the polypeptide in a sufficient concentration to modulate the activity of the polypeptide.

In another embodiment, the invention provides a method of inhibiting aberrant activity of a SLC5A8-expressing cell, comprising contacting the cell with a compound that modulates the activity or expression of the polypeptide, in an amount which is effective to reduce or inhibit the aberrant activity of the cell.

In certain embodiments, compounds used in the methods of the invention are selected from the group consisting of a peptide, a phosphopeptide, a small organic molecule, an antibody, and a peptidomimetic. Cells in the methods of the invention can be found in the colon, kidney, lung, esophagus, small bowel, stomach, thyroid, uterus, and breast.

In another embodiment, the invention provides a method of treating or preventing a disorder characterized by aberrant activity of a SLC5A8-expressing cell, in a subject, comprising administering to the subject an effective amount of a compound that modulates the activity or expression of the SLC5A8 polypeptide, such that the aberrant activity of the SLC5A8-expressing cell is reduced or inhibited.

In another embodiment, the invention provides a transgenic mouse having germline and somatic cells comprising a chromosomally incorporated transgene that disrupts the genomic SLC5A8 gene and inhibits expression of said gene, wherein said disruption comprises insertion of a selectable marker sequence resulting in said transgenic mouse exhibiting increased susceptibility to the formation of tumors as compared to the wildtype mouse. The transgenic mouse can be homozygous r heterozygous for the disruption.

In another embodiment, the invention provides a transgenic mouse having germline and somatic cells in which at least one allele of a genomic SLC5A8 gene is disrupted by a chromosomally incorporated transgene, which transgene inhibits the expression of the genomic SLC5A8 gene, wherein (i) the genomic SLC5A8 gene encodes a SLC5A8 protein; and (ii) the disruption comprises insertion of a selectable marker sequence, which replaces all or a portion of the genomic SLC5A8 gene or is inserted into the coding sequence of the genomic SLC5A8 gene; and (iii) the transgenic mouse has increased susceptibility to the development of neoplasms.

In another embodiment, the invention provides isolated mammalian cells comprising a diploid genome including a chromosomally incorporated transgene, which transgene disrupts the genomic SLC5A8 gene and inhibits expression of said gene. Optionally, the cells are mouse cells.

In another embodiment, the invention provides a method for generating a mouse and mouse embryonic stem cells having a functionally disrupted endogenous SLC5A8 gene, comprising the steps of: (i) constructing a transgene construct including (a) a recombination region having all or a portion of the endogenous SLC5A8 gene, which recombination region directs recombination of the transgene with the endogenous SLC5A8 gene; and (b) a marker sequence which provides a detectable signal for identifying the presence of the transgene in a cell; (ii) transferring the transgene into embryonic stem cells of a mouse; (iii) selecting embryonic stem cells having a correctly targeted homologous recombination between the transgene and the SLC5A8 gene; (iv) transferring said cells identified in step (iii) into a mouse blastocyst and implanting the resulting chimeric blastocyst into a female mouse; and (v) selecting offspring harboring an endogenous SLC5A8 gene allele comprising the correctly targeted recombination.

In another embodiment, the invention provides a method of evaluating the carcinogenic potential of an agent comprising: (i) contacting the transgenic mouse of claim 16A with a test agent; and (ii) comparing the number of transformed cells in a sample from the treated mouse with the number of transformed cells in a sample from an untreated transgenic mouse or transgenic mouse treated with a control agent, wherein the difference in the number of transformed cells in the treated mouse, relative to the number of transformed cells in the absence of treatment or treatment with a control agent, indicates the carcinogenic potential of the test compound.

In another embodiment, the invention provides a method of evaluating an anti-proliferative activity of a test compound, comprising: (i) providing a transgenic mouse of claim 16A having germline and somatic cells in which the expression of the SLC5A8 gene is inhibited by said chromosomally incorporated transgene, or a sample of cells derived therefrom; (ii) contacting the transgenic mouse or the sample of cells with a test agent; and (iii) determining the number of transformed cells in a specimen from the transgenic mouse or in the sample of cells, wherein a statistically significant decrease in the number of transformed cells, relative to the

number of transformed cells in the absence of the test agent, indicates the test compound is a potential anti-proliferative agent.

In certain aspects, the present invention is based, at least in part, on Applicants' discovery of a particular human genomic DNA region in which the cytosines within CpG dinucleotides are methylated in tissues from human cancers and unmethylated in normal human tissues. The region is referred to hereinafter as the "SLC5A8-methylation target region" is encompassed by base pairs 82200 to 83267 of GenBank entry AC063951, and is located in the promoter and/or exon 1 of the SLC5A8 gene. The present methods are also based, at least in part, on Applicants' discovery that the levels of SLC5A8 transcript in tissues from human cancers are lower than the levels of SLC5A8 transcript in normal tissues.

In one embodiment, the method comprises assaying for the presence of differentially methylated SLC5A8 nucleotide sequences (e.g., in the SLC5A8 methylation target region) in a tissue sample or a bodily fluid sample from a subject. Preferred bodily fluids include blood, serum, plasma, a blood-derived fraction, stool, colonic effluent or urine. In one embodiment, the method involves restriction enzyme/methylation-sensitive PCR. In another embodiment, the method comprises reacting DNA from the sample with a chemical compound that converts non-methylated cytosine bases (also called "conversion-sensitive" cytosines), but not methylated cytosine bases, to a different nucleotide base. In a preferred embodiment, the chemical compound is sodium bisulfite, which converts unmethylated cytosine bases to uracil. The compound-converted DNA is then amplified using a methylation-sensitive polymerase chain reaction (MSP) employing primers that amplify the compound-converted DNA template if cytosine bases within CpG dinucleotides of the DNA from the sample are methylated. Production of a PCR product indicates that the subject has cancer or precancerous adenomas. Other methods for assaying for the presence of methylated DNA are known in the art.

In another embodiment, the method comprises assaying for decreased levels of an SLC5A8 transcript in the sample. A sequence of the SLC5A8 transcript (SEQ ID NO: 3) is shown in Figure 2. The SLC5A8 transcript is encoded by 15 exons within the present genomic contig. In another aspect the method comprises assaying for decreased levels of a protein encoded by the SLC5A8 transcript in the sample.

In another embodiment, the present invention provides a detection method for prognosis of a cancer (e.g., colon cancer, breast cancer, thyroid cancer, or stomach cancer) in a subject

known to have or suspected of having cancer. Such method comprises assaying for the presence of methylated SLC5A8 DNA (e.g., in the SLC5A8 methylation target region) in a tissue sample or bodily fluid from the subject. In certain cases, it is expected that detection of methylated SLC5A8 DNA in a blood fraction is indicative of an advanced state of cancer (e.g., colon cancer). In other cased, detection of methylated SLC5A8 DNA in a tissue or stool derived sample or sample from other bodily fluids may be indicative of a cancer that will respond to therapeutic agents that demethylate DNA or reactivate expression of the SLC5A8 gene.

In another embodiment, the present invention provides a method for monitoring over time the status of cancer (e.g., colon cancer, breast cancer, thyroid cancer, or stomach cancer) in a subject. The method comprises assaying for the presence of methylated SLC5A8 DNA (e.g., in the SLC5A8 methylation target region) in a tissue sample or bodily fluid taken from the subject at a first time and in a corresponding tissue sample or bodily fluid taken from the subject at a second time. Absence of methylated SLC5A8 DNA from the tissue sample or bodily fluid taken at the first time and presence of methylated SLC5A8 DNA in the tissue sample or bodily fluid taken at the second time indicates that the cancer is progressing. Presence of methylated SLC5A8 DNA in the tissue sample or bodily fluid taken at the first time and absence of methylated SLC5A8 DNA from the tissue sample or bodily fluid taken at the second time indicates that the cancer is regressing.

In another embodiment, the present invention provides a method for evaluating therapy in a subject having cancer or suspected of having cancer (e.g., colon cancer, breast cancer, thyroid cancer, or stomach cancer). The method comprises assaying for the presence of methylated SLC5A8 DNA (e.g., in the SLC5A8 methylation target region) in a tissue sample or bodily fluid taken from the subject prior to therapy and a corresponding bodily fluid taken from the subject during or following therapy. Loss of or a decrease in the levels of methylated SLC5A8 DNA in the sample taken after or during therapy as compared to the levels of methylated SLC5A8 DNA in the sample taken before therapy is indicative of a positive effect of the therapy on cancer regression in the treated subject.

The present invention also relates to oligonucleotide primer sequences for use in assays (e.g., methylation-sensitive PCR assays or HpaII assays) designed to detect the methylation status of the SLC5A8 gene. The present invention also relates to antibodies and to oligonucleotides or oligomers for detecting the presence the SLC5A8 protein or the SLC5A8 [illegible] respectively, in samples obtained from a subject.

The present invention also provides a method of inhibiting or reducing growth of cancer cells (e.g., colon cancer, breast cancer, thyroid cancer, or stomach cancer). The method comprises increasing the levels of the protein encoded by SLC5A8 in cancer cells. In one embodiment, the cells are contacted with the SLC5A8 protein or a biologically active equivalent or fragment thereof under conditions permitting uptake of the protein or fragment. In another embodiment, the cells are contacted with a nucleic acid encoding the SLC5A8 protein and comprising a promoter active in the cancer cell, wherein the promoter is operably linked to the region encoding the SLC5A8 protein, under conditions permitting the uptake of the nucleic acid by the cancer cell. In another embodiment, the method comprises demethylating the methylated SLC5A8 DNA, or otherwise reactivating the silenced SLC5A8 promoter.

In one embodiment, the application provides isolated or recombinant SLC5A8 nucleotide sequences that are at least 80%, 85%, 90%, 95%, 98%, 99% or identical to the nucleotide sequence of any one of SEQ ID NOs: 2-4 and 21, fragments of said sequences that are 10, 15, 20, 25, 50, 100, or 150 base pairs in length wherein the SLC5A8 nucleotide sequences are differentially methylated in an SLC5A8-associated disease cell.

In another embodiment, the application provides a method for detecting colon cancer, comprising: a) obtaining a sample from a patient; and b) assaying said sample for the presence of methylation of nucleotide sequences within at least two genes selected from the group consisting of: SLC5A8, HLTF, p16, and hMLH1; wherein methylation of nucleotide sequences within the two genes is indicative of colon cancer. In such methods, the sample is a bodily fluid selected from the group consisting of blood, serum, plasma, a blood-derived fraction, stool, urine, and a colonic effluent. For example, the bodily fluid is obtained from a subject suspected of having or is known to have colon cancer.

In another embodiment, the application provides a kit for detecting colon cancer in a subject, comprising primers for detecting methylation of nucleotide sequence within at least two genes selected from the group consisting of: SLC5A8, HLTF, p16, and hMLH1, wherein the primers for detecting methylation of SLC5A8 nucleotide sequence are selected from SEQ ID NOs: 5-11; wherein the primers for detecting methylation of HLTF nucleotide sequence are selected from 5'-TGGGGTTTCGTGGTTTTTTCGCGC-3', 5'-CCGCGAATCCAATCAAACGTCGACG-3', 5'-ATTTTTGGGGTTTTGTGGTTTTTTTGTGT-3', 5'-ATCAACCACAAATCCAATCAAACATCAACA-3', 5'-

GCACGACTAAAAAATAAATCGCCGCG-3', 5'-AAACACACAACTAAAAAATAAATCACCACA-3', 5'-TAAAACCTCGTAACTTTCCCGCGCG-3', 5'-GTCGCGAGTTTAGTTAGACGTCGAC-3', 5'-TCCTAAAACCTCATAACTTTCCCACACA-3', and 5'-AGTTGTTGTGAGTTTAGTTAGATGTTGAT-3', wherein the primers for detecting methylation of hMLH1 nucleotide sequence are selected from 5'AACGAATTAATAGGAAGAGCGGATAGCG-3', 5'-CGTCCCTCCCTAAAACGACTACTACCC-3', 5'-CGTTTTTTTTTGAAGCGGTTATTGTTTGT-3', and 5'-AACGAACCAATAAAAAAAACAAACAACG-3'. Tthe kit may further comprise a compound to convert a template DNA. Optioanally the compound is bisulfite.

Brief Description of The Drawings

Figure 1 shows the complete sequence of the Genomic clone AC063951 (SEQ ID NO: 2), with nucleotides 82200-83267 underlined on pages 35 of Figure 1. This region (nucleotides 82200-83267 of AC063951, SEQ ID NO: 12, see Figure 4) encompasses the promoter and/or exon 1 of the SLC5A8 gene, and is herein referred to as the "SLC5A8 methylation target region."

Figure 2 shows the nucleotide sequence of the SLC5A8 mRNA transcript (SEQ ID NO: 3). The SLC5A8 transcript is encoded by 15 exons within the present genomic contig.

Figure 3 shows a diagram of the SLC5A8 methylation target region. CpG sites are shown with circles and stems. The numerical coordinates are those of genomic clone AC063951. Lollipops designate CpG sites that are potential acceptors of aberrant methylation. Asterisks designate sites recognized by the HpaII restriction enzyme. Shown are the positions of PCR primers that amplify regions crossing 6 HpaII sites, or regions crossing 4 HpaII sites. Also shown is the position of PCR primers designed for a methyl-specific PCR (MS-PCR) assays. Also shown in the gray bar is the 5' end of exon 1 of the SLC5A8 transcript which overlaps with the methylation sites detected in both MS-PCR and HpaII based assays. Lastly indicated is a NotI site corresponding to methylation site 2D41 detected in Restriction Landmark Genome Scanning assay as methylated in colon cancer cell lines, though not in primary tumors.

Figure 4 provides the sequence of AC063951 between nucleotides 82200-83267 (SEQ ID NO: 12), and designates every CpG site with a gray lollipop, and shows the HpaII sites in the assay as dark lollipops, and also shows the location of the PCR primers used in the assay. In this figure, the base pairs have been renumbered sequentially from 1-1068, with nucleotide 82200 being renumbered as nucleotide 1.

Figure 5 shows the correlation between HpaII assays (over 4 HpaII sites and 6 HpaII sites) and silencing of expression of the SLC5A8 transcript.

Figure 6 shows the results of the HpaII assays (over 4 HpaII sites and 6 HpaII sites) in actual colon cancer tumors and normal control colon tissues.

Figure 7 shows the results of assay for methylation at 61 CpG sites enumerated in Figure 4 with site 1 corresponding to basepair 466 in Figure 4 and site 61 corresponding to basepair 1010. The bold arrows correspond to 4 of the HpaII sites at respectively basepairs 466, 691, 709, and 716 in Figure 4. Methylation was assayed by sequencing DNA from samples following sodium bisulfite treatment of DNA that converts cytosine to uracil but leaves methyl-cytosine unchanged. Bases that are methylated are coded black, unmethylated bases are coded dark gray, and samples with both methylated and unmethylated bases are coded light gray.

Figure 8 shows the wild-type sequence of the anti-sense strand of AC063951 between bases 82200-83267 (SEQ ID NO: 13). Note that the sequence is the reverse complement of that shown in Figure 4, and therefore base number 1 on this diagram corresponds to basepair 83267 in AC063951, and to basepair 1068 in Figure 4. Indicated on this diagram is the position of the MS-PCR1 primers (AS-meth) and the UMS-PCR1 primers (AS-unmethy). The methyl specific MS-PCR1 primers amplify a CpG sites numbered 6, 7, 8 and 15, 16, 17, 18 respectively in Figure 7. The UMS-PCR1 primers interrogate CpG sites 7, 8 and 15, 16, 17, 18 respectively.

Figure 9 shows a region within SEQ ID NO: 13 shown in Figure 8 (nucleotides 300-600, SEQ ID NO: 14), and the sequences of the antisense strand that are amplified by the methyl-specific and unmethyl-specific PCR primers.

Figure 10 shows the bisulfite converted sequence of a uniformly methylated SLC5A8 antisense strand (SEQ ID NO: 15), but not the wild-type sequence of the SLC5A8 antisense strand (corresponding to Figure 8). Indicated again are the position of the methylation specific PCR primers for the MS-PCR1 assay.

Figure 11 shows the bisulfite converted sequence of a uniformly unmethylated SLC5A8 antisense strand (SEQ ID NO: 16), but not the wild-type sequence of the SLC5A8 antisense strand shown in Figure 8. Indicated are the position of the unmethylation specific PCR primers for the UMS-PCR1 assay.

Figure 12 provides the bisulfite converted sequence of the unmethylated SLC5A8 sense strand of nucleotides 82200-83267 of AC063951, renumbered such that basepair 82200 is designated as nucleotide 1 (SEQ ID NO: 17).

Figure 13 provides the bisulfite converted sequence of a uniformly methylated SLC5A8 sense strand of nucleotides 82200-83267 (SEQ ID NO: 18).

Figure 14 shows the tabular results of MS-PCR1 assay performed on 31 colon cancer cell lines that do or do not express the SLC5A8 transcript.

Figure 15 shows the tabular results of MS-PCR1 assay performed on 63 matched sets of primary colon cancer tumor tissue and accompanying normal colon tissue.

Figure 16 shows the results of testing 12 normal colon tissues from individuals without colon cancer.

Figure 17 shows the tabular results of the MS-PCR1 assay of 28 premalignant colon adenomas, 68% of which are detected.

Figure 18 shows the amino acid sequence (SEQ ID NO: 1) of the SLC5A8 protein.

Figure 19 shows RT-PCR detection of the SLC5A8 transcript in normal colon and in a minority subset of colon cancer cell lines.

Figure 20 shows RT-PCR detection of SLC5A8 transcript in colon cancer cell lines that have been treated with the DNA-demethylating agent 5-azacytidine. 5-azacytidine reactivates expression of the SLC5A8 gene in 6 of 8 colon cancer cell lines.

Figure 21 demonstrates detection of methylation of the SLC5A8 locus by showing resistance of the locus to HpaII digestion. The 4 HpaII assay (as described in the invention disclosure) is based on PCR amplification of a portion of the SLC5A8 locus. Lanes labeled U show control amplification of undigested SLC5A8 DNA. Lanes labeled M show amplification [illegible]DNA[illegible] that has first been cut with the restriction enzyme Msp1.

Figure 22 demonstrates detection of SLC5A8 DNA methylation in primary colon cancer tumors but not in matched normal tissue from the same patients. Samples labeled T represent colon cancer tumor tissue; whereas samples labeled N represent the matched normal tissue.

Figures 23A-23B show the identification of SLC5A8. (A) Shown is the genomic structure of the SLC5A8 gene. Black boxes represent exons, and arrows the start codon and stop codons respectively. (B) The nucleotide sequence of the SLC5A8 coding region (SEQ ID NO: 4).

Figures 24A-24F show SLC5A8 expression. (A) Shown is RT-PCR analysis demonstrating *SLC5A8* transcript expression in three normal colon mucosa samples (N1, N2, N3), but absence of *SLC5A8* transcript in most colon cancer cell lines (remaining samples). (B) Shown is RT-PCR analysis demonstrating reactivation of SLC5A8 expression in cell lines treated with 5-azacytidine (+) compared to untreated (-) controls. (C) Methylation specific PCR (MS-PCR) assay for methylated (M) or unmethylated (U) *SLC5A8* exon 1 sequences detects exclusively methylated templates in SLC5A8 silenced cell lines. (D) MS-PCR detects only unmethylated *SLC5A8* templates in SLC5A8 expressing cell lines. (E) MS-PCR detection of methylated *SLC5A8* templates in colon cancer tumors (T) antecedent to *SLC5A8* methylated cell lines (V425, V670). Matched normal colon tissue (N) shows only unmethylated templates. Unmethylated templates in tumor tissue presumptively arise from contaminating non-malignant cells. (F) MS-PCR analysis of colon cancer tumors (T) and matched normal (N) colon tissues. Methyl specific bands are seen in each of the tumor samples, but none of the normal controls.

Figures 25A-25B show real time MS-PCR analysis of *SLC5A8* methylation. Plotted are 1000 times the ratio of measured *SLC5A8* methylated product to the control *MYOD1* derived product. (A) Detection of *SLC5A8* methylation in primary colon cancer tissues. Column 1 displays values for normal colon tissues harvested from non-cancer resections (dark diamonds). Column 2 displays values for normal colon tissues harvested from colon cancer resections (dark diamonds). Column 3 displays values for colon cancer tissues divided into unmethylated samples falling within the normal tissue range (dark diamonds at the bottom), versus methylated samples showing values greater than the normal tissue range (light diamonds at the top). Adjacent bars indicate population means. (B) Real time MS-PCR analysis of *SLC5A8* methylation in aberrant crypt foci. Column 1 displays values for 24 normal colon tissues harvested from colon resections from 11 individuals (dark diamonds). Column 2 displays values for 15 aberrant crypt foci harvested from the same 11 individuals' resections. Dark diamonds (at

the bottom) indicate unmethylated samples within the normal range, and light diamonds (at the top) indicate methylated samples falling within the range previously demonstrated by methylated cancers. Adjacent bars indicate the mean value for each group.

Figure 26 shows real time MS-PCR analysis of *SLC5A8* methylation in DNA precipitated from the serum of colon cancer patients. Plotted are 1000 times the ratio of measured *SLC5A8* methylated product to the control *MYOD1* derived product. Column 1 displays absence of detectable *SLC5A8* methylation in serum of 13 individuals whose colon cancer tumors assayed as unmethylated by MS-PCR (dark diamonds at the bottom). Column 2 displays values of *SLC5A8* methylation in the serum of 10 individuals whose colon cancer tumors assayed as methylated by MS-PCR. Dark diamonds (at the bottom) indicate 6 sera without detectable methylation, and light diamonds (at the top) indicate 4 sera in which *SLC5A8* methylation was detectable.

Figures 27A-27B show SLC5A8 suppression of colon cancer colony formation. Shown are the number of G418 resistant colonies arising from transfection with a SLC5A8 expression vector (SLC5A8) or a control empty expression vector (pcDNA) in *SLC5A8* unmethylated and expressing V364, V457, and V9M cells (panel A) as compared to *SLC5A8* methylated and deficient FET, V400, and RKO cells (panel B).

Figure 27 shows the cloning of *SLC5A8* transcript. Black bars indicate representative ESTs. The lighter gray bar indicates sequence generated from an image clone. The dark gray bar indicates open reading frame encoding SLC5A8 protein.

Figure 28 shows the protein alignments of SLC5A8, the closest murine homologue of SLC5A8, the human sodium iodide symporter SLC5A5, and the human sodium dependent multivitamin transporter SLC5A6.

Figures 30A-30B show methylation in *SLC5A8* exon 1. (A) Diagrammatic representation of the CpG island in *SLC5A8* exon 1. Balloons represent CpG dinucleotides. Coordinates represent nucleotide positions numbered as per GenBank entry AC063951. Positions of the ATG and NotI site are indicated. Arrows cover the regions interrogated by primers for MS-PCR. (B) Diagrammatic summary of methylation status of the 62 CpG sites in *SLC5A8* exon 1 as determined by sequencing of bisulfite converted genomic DNA. Each site is sequentially represented by one shaded block. Black represents sites that are fully methylated. / represents sites that are fully unmethylated. And lighter gray represents sites that

are partially methylated. Samples include 9 SLC5A8 silenced cell lines (Off samples), 6 SLC5A8 expressing normal colonic mucosa (On samples designated N), and 3 SLC5A8 expressing cell lines (On samples designated V). Arrows indicate sites that are interrogated by MS-PCR primers and bracket a differentially methylated region that is unmethylated in SLC5A8 expressing samples and is methylated in SLC5A8 silenced samples.

Figure 30 shows methylation events in primary colon cancers. Shown is analysis of 64 primary colon cancers for aberrant methylation at 4 genomic loci, SLC5A8, HLTF, hMLH1, and p16. Black bars represent positive assays for methylation in tumor tissue, and gray bars represent detection only of unmethylated alleles.

Figure 31 shows suppression of xenograft growth in 4 of 5 SLC5A8 expressing V400 transfected clones (square symbols, gray lines) as compared with control pools of V400 cells transfected with an empty expression vector (triangular symbols, black lines).

Detailed Description of the Invention

I. Definitions

For convenience, certain terms employed in the specification, examples, and appended claims are collected here. Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs.

The articles "a" and "an" are used herein to refer to one or to more than one (i.e., to at least one) of the grammatical object of the article, unless the context clearly indicates otherwise. By way of example, "an element" means one element or more than one element.

The terms "adenoma", "colon adenoma," and "polyp" are used herein to describe any precancerous neoplasia of the colon.

The term "blood-derived fraction" herein refers to a component or components of whole blood. Whole blood comprises a liquid portion (i.e., plasma) and a solid portion (i.e., blood cells). The liquid and solid portions of blood are each comprised of multiple components; e.g., different proteins in plasma or different cell types in the solid portion. One of these components or a mixture of any of these components is a blood-derived fraction as long as such fraction is missing one or more components found in whole blood.

"Cells," "host cells" or "recombinant host cells" are terms used interchangeably herein. It is understood that such terms refer not only to the particular subject cell but to the progeny or potential progeny of such a cell. Because certain modifications may occur in succeeding generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term as used herein.

A "chimeric polypeptide" or "fusion polypeptide" is a fusion of a first amino acid sequence with a second amino acid sequence where the first and second amino acid sequences are not naturally present in a single polypeptide chain.

The term "colon" as used herein is intended to encompass the right colon (including the cecum), the transverse colon, the left colon, and the rectum.

The terms "colorectal cancer" and "colon cancer" are used interchangeably herein to refer to any cancerous neoplasia of the colon (including the rectum, as defined above).

The terms "compound", "test compound," and "agent" are used herein interchangeably and are meant to include, but are not limited to, peptides, nucleic acids, carbohydrates, small organic molecules, natural product extract libraries, and any other molecules (including, but not limited to, chemicals, metals, and organometallic compounds).

The term "compound-converted DNA" herein refers to DNA that has been treated or reacted with a chemical compound that converts unmethylated C bases in DNA to a different nucleotide base. For example, one such compound is sodium bisulfite, which converts unmethylated C to U. If DNA that contains conversion-sensitive cytosine is treated with sodium bisulfite, the compound-converted DNA will contain U in place of C. If the DNA which is treated with sodium bisulfite contains only methylcytosine, the compound-converted DNA will not contain uracil in place of the methylcytosine.

The term "de-methylating agent" as used herein refers agents that restore activity and/or gene expression of target genes silenced by methylation upon treatment with the agent. Examples of such agents include without limitation 5-azacytidine and 5-aza-2'-deoxycytidine.

The term "detection" is used herein to refer to any process of observing a marker, in a biological sample, whether or not the marker is actually detected. In other words, the act of probing a sample for a marker is a "detection" even if the marker is determined to be not present

or below the level of sensitivity. Detection may be a quantitative, semi-quantitative or non-quantitative observation.

The term "differentially methylated SLC5A8 nucleotide sequence" refers to a region of the SLC5A8 nucleotide sequence that is found to be methylated in a SLC5A8-associated cancer such as a region of the SLC5A8 nucleotide sequence that is found to be methylated in cancer tissues or cell lines, but not methylated in the normal tissues or cell lines. For example, Figure 3 delineates certain SLC5A8 regions that are differentially methylated, such as SEQ ID NOs: 11-13.

"Expression vector" refers to a replicable DNA construct used to express DNA which encodes the desired protein and which includes a transcriptional unit comprising an assembly of (1) genetic element(s) having a regulatory role in gene expression, for example, promoters, operators, or enhancers, operatively linked to (2) a DNA sequence encoding a desired protein (in this case, a SLC5A8 protein) which is transcribed into mRNA and translated into protein, and (3) appropriate transcription and translation initiation and termination sequences. The choice of promoter and other regulatory elements generally varies according to the intended host cell. In general, expression vectors of utility in recombinant DNA techniques are often in the form of "plasmids" which refer to circular double stranded DNA loops which, in their vector form are not bound to the chromosome. In the present specification, "plasmid" and "vector" are used interchangeably as the plasmid is the most commonly used form of vector. However, the invention is intended to include such other forms of expression vectors which serve equivalent functions and which become known in the art subsequently hereto.

In the expression vectors, regulatory elements controlling transcription or translation can be generally derived from mammalian, microbial, viral or insect genes. The ability to replicate in a host, usually conferred by an origin of replication, and a selection gene to facilitate recognition of transformants may additionally be incorporated. Vectors derived from viruses, such as retroviruses, adenoviruses, and the like, may be employed.

As used herein, the phrase "gene expression" or "protein expression" includes any information pertaining to the amount of gene transcript or protein present in a sample, as well as information about the rate at which genes or proteins are produced or are accumulating or being degraded (e.g., reporter gene data, data from nuclear runoff experiments, pulse-chase data etc.). Certain kinds of data might be viewed as relating to both gene and protein expression. For

example, protein levels in a cell are reflective of the level of protein as well as the level of transcription, and such data is intended to be included by the phrase "gene or protein expression information." Such information may be given in the form of amounts per cell, amounts relative to a control gene or protein, in unitless measures, etc.; the term "information" is not to be limited to any particular means of representation and is intended to mean any representation that provides relevant information. The term "expression levels" refers to a quantity reflected in or derivable from the gene or protein expression data, whether the data is directed to gene transcript accumulation or protein accumulation or protein synthesis rates, etc.

The terms "healthy", "normal," and "non-neoplastic" are used interchangeably herein to refer to a subject or particular cell or tissue that is devoid (at least to the limit of detection) of a disease condition, such as a neoplasia (e.g., cancer), that is associated with SLC5A8 such as for example neoplasia associated with silencing of SLC5A8 gene expression due to methylation. These terms are often used herein in reference to tissues and cells of the colon. Thus, for the purposes of this application, a patient with severe heart disease but lacking a SLC5A8 silencing-associated disease would be termed "healthy."

"Homology" or "identity" or "similarity" refers to sequence similarity between two peptides or between two nucleic acid molecules. Homology and identity can each be determined by comparing a position in each sequence which may be aligned for purposes of comparison. When an equivalent position in the compared sequences is occupied by the same base or amino acid, then the molecules are identical at that position; when the equivalent site occupied by the same or a similar amino acid residue (e.g., similar in steric and/or electronic nature), then the molecules can be referred to as homologous (similar) at that position. Expression as a percentage of homology/similarity or identity refers to a function of the number of identical or similar amino acids at positions shared by the compared sequences. A sequence which is "unrelated" or "non-homologous" shares less than 40% identity, preferably less than 25% identity with a sequence of the present invention. In comparing two sequences, the absence of residues (amino acids or nucleic acids) or presence of extra residues also decreases the identity and homology/similarity.

The term "homology" describes a mathematically based comparison of sequence similarities which is used to identify genes or proteins with similar functions or motifs. The nucleic acid and protein sequences of the present invention may be used as a "query sequence" to perform a search against public databases to, for example, identify other family members,

related sequences or homologs. Such searches can be performed using the NBLAST and XBLAST programs (version 2.0) of Altschul, et al. (1990) *J Mol. Biol.* 215:403-10. BLAST nucleotide searches can be performed with the NBLAST program, score=100, wordlength=12 to obtain nucleotide sequences homologous to nucleic acid molecules of the invention. BLAST protein searches can be performed with the XBLAST program, score=50, wordlength=3 to obtain amino acid sequences homologous to protein molecules of the invention. To obtain gapped alignments for comparison purposes, Gapped BLAST can be utilized as described in Altschul et al., (1997) Nucleic Acids Res. 25(17):3389-3402. When utilizing BLAST and Gapped BLAST programs, the default parameters of the respective programs (e.g., XBLAST and BLAST) can be used. See http://www.ncbi.nlm.nih.gov.

As used herein, "identity" means the percentage of identical nucleotide or amino acid residues at corresponding positions in two or more sequences when the sequences are aligned to maximize sequence matching, i.e., taking into account gaps and insertions. Identity can be readily calculated by known methods, including but not limited to those described in (*Computational Molecular Biology*, Lesk, A. M., ed., Oxford University Press, New York, 1988; *Biocomputing*: Informatics and Genome Projects, Smith, D. W., ed., Academic Press, New York, 1993; *Computer Analysis of Sequence Data*, Part I, Griffin, A. M., and Griffin, H. G., eds., Humana Press, New Jersey, 1994; *Sequence Analysis in Molecular Biology*, von Heinje, G., Academic Press, 1987; and *Sequence Analysis Primer*, Gribskov, M. and Devereux, J., eds., M Stockton Press, New York, 1991; and Carillo, H., and Lipman, D., SIAM J. Applied Math., 48: 1073, 1988). Methods to determine identity are designed to give the largest match between the sequences tested. Moreover, methods to determine identity are codified in publicly available computer programs. Computer program methods to determine identity between two sequences include, but are not limited to, the GCG program package (Devereux, J., et al., *Nucleic Acids Research* 12(1): 387 (1984)), BLASTP, BLASTN, and FASTA (Altschul, S. F. et al., *J. Molec. Biol.* 215: 403-410 (1990) and Altschul et al. *Nuc. Acids Res.* 25: 3389-3402 (1997)). The BLAST X program is publicly available from NCBI and other sources (*BLAST Manual*, Altschul, S., et al., NCBI NLM NIH Bethesda, Md. 20894; Altschul, S., et al., *J. Mol. Biol.* 215: 403-410 (1990)). The well known Smith Waterman algorithm may also be used to determine identity.

"SLC5A8-associated cancer" refers to cancer associated with reduced expression or no expression of the SLC5A8 gene (previously referred to as the Hui1 gene), and cancer associated with aberrant methylation of SLC5A8 DNA. Examples of SLC5A8-associated cancer

include, but are not limited to, colon cancer, breast cancer, thyroid cancer, and stomach cancer. As used herein, the SLC5A8-associated cancers includes both cancers and pre-cancer adenomas.

"SLC5A8-associated proliferative disorder" refers to a disease that is associated with either reduced expression or over-expression of the SLC5A8 gene.

A "SLC5A8-associated protein" refers to a protein capable of interacting with and/or binding to a SLC5A8 polypeptide. Generally, the SLC5A8-associated protein may interact directly or indirectly with the SLC5A8 polypeptide.

"SLC5A8-methylation target regions" as used herein refer to those regions of SLC5A8 that are found to be methylated. These regions include nucleotide regions that may be either constitutively or differentially methylated regions. For example, Figure 3 discloses a SLC5A8 region wherein certain sequences of this region are differentially methylated regions.

"SLC5A8-nucleotide sequence" or "SLC5A8-nucleic acid sequence" as used herein refers to the SLC5A8 nucleotide sequences as set forth in SEQ ID NOs: 2-7 and fragments thereof.

"SLC5A8-silencing associated diseases" as used herein includes SLC5A8-associated cancer.

The term "including" is used herein to mean, and is used interchangeably with, the phrase "including but not limited to."

The term "isolated" as used in reference to nucleic acids or polypeptides indicates a nucleic acid or polypeptide, such as a SLC5A8 nucleic acid or polypeptide, that is isolated from, or otherwise substantially free of other proteins that are normally associated with the nucleic acid or polypeptide.

The term "methylation-sensitive PCR" (i.e., MSP) herein refers to a polymerase chain reaction in which amplification of the compound-converted template sequence is performed. Two sets of primers are designed for use in MSP. Each set of primers comprises a forward primer and a reverse primer. One set of primers, called methylation-specific primers, will amplify the compound-converted template sequence if C bases in CpG dinucleotides within the template DNA (e.g., a SLC5A8 nucleic acid) are methylated. Another set of primers, called unmethylation-specific primers, will amplify the compound-converted template sequences if C

bases in CpG dinucleotides within the template DNA (e.g., a SLC5A8 nucleic acid) are not methylated.

The term "nucleic acid" refers to polynucleotides such as deoxyribonucleic acid (DNA), and, where appropriate, ribonucleic acid (RNA). The term should also be understood to include, as equivalents, analogs of either RNA or DNA made from nucleotide analogs, and, as applicable to the embodiment being described, single (sense or antisense) and double-stranded polynucleotides.

"Operably linked" when describing the relationship between two DNA regions simply means that they are functionally related to each other. For example, a promoter or other transcriptional regulatory sequence is operably linked to a coding sequence if it controls the transcription of the coding sequence.

The term "or" is used herein to mean, and is used interchangeably with, the term "and/or", unless context clearly indicates otherwise.

The terms "polypeptide" and "protein" are used interchangeably herein.

The term "recombinant" as used in reference to a nucleic acid indicates any nucleic acid that is positioned adjacent to one or more nucleic acid sequences that it is not found adjacent to in nature. A recombinant nucleic acid may be generated in vitro, for example by using the methods of molecular biology, or in vivo, for example by insertion of a nucleic acid at a novel chromosomal location by homologous or non-homologous recombination. The term "recombinant" as used in reference to a polypeptide indicates any polypeptide that is produced by expression and translation of a recombinant nucleic acid.

A "sample" includes any material that is obtained or prepared for detection of a molecular marker or a change in a molecular marker such as the methylation state, or any material that is contacted with a detection reagent or detection device for the purpose of detecting a molecular marker or a change in the molecular marker.

A "subject" is any organism of interest, generally a mammalian subject, such as a mouse, and preferably a human subject.

The term "transgene" is used herein to describe genetic material which has been or is about to be artificially inserted into the genome of a mammal, particularly a mammalian cell of a ıal. By "transgenic animal" is meant a non-human animal, usually a mammal (e.g.,

mouse, rat, rabbit, hamster, etc.), having a non-endogenous nucleic acid sequence present as an extrachromosomal element in a portion of its cells or stably integrated into its germ line DNA (i.e., in the genomic sequence of most or all of its cells). Heterologous nucleic acid is introduced into the germ line of such transgenic animals by genetic manipulation of, for example, embryos or embryonic stem cells of the host animal.

II. Overview

In certain aspects, the invention relates, in part, to methods for determining whether a patient is likely or unlikely to have a cancer, for example, colon neoplasia. A colon neoplasia is any cancerous or precancerous growth located in, or derived from, the colon. The colon is a portion of the intestinal tract that is roughly three feet in length, stretching from the end of the small intestine to the rectum. Viewed in cross section, the colon consists of four distinguishable layers arranged in concentric rings surrounding an interior space, termed the lumen, through which digested materials pass. In order, moving outward from the lumen, the layers are termed the mucosa, the submucosa, the muscularis propria and the subserosa. The mucosa includes the epithelial layer (cells adjacent to the lumen), the basement membrane, the lamina propria and the muscularis mucosae. In general, the "wall" of the colon is intended to refer to the submucosa and the layers outside of the submucosa. The "lining" is the mucosa.

Precancerous colon neoplasias are referred to as adenomas or adenomatous polyps. Adenomas are typically small mushroom-like or wart-like growths on the lining of the colon and do not invade into the wall of the colon. Adenomas may be visualized through a device such as a colonoscope or flexible sigmoidoscope. Several studies have shown that patients who undergo screening for and removal of adenomas have a decreased rate of mortality from colon cancer. For this and other reasons, it is generally accepted that adenomas are an obligate precursor for the vast majority of colon cancers. When a colon neoplasia invades into the basement membrane of the colon, it is considered a colon cancer, as the term "colon cancer" is used herein. In describing colon cancers, this specification will generally follow the so-called "Dukes" colon cancer staging system. The characteristics that the describe a cancer are generally of greater significance than the particular term used to describe a recognizable stage. The most widely used staging systems generally use at least one of the following characteristics for staging: the extent of tumor penetration into the colon wall, with greater penetration generally correlating with a more dangerous tumor; the extent of invasion of the tumor through the colon wall and into other neighboring tissues, with greater invasion generally correlating

with a more dangerous tumor; the extent of invasion of the tumor into the regional lymph nodes, with greater invasion generally correlating with a more dangerous tumor; and the extent of metastatic invasion into more distant tissues, such as the liver, with greater metastatic invasion generally correlating with a more dangerous disease state.

"Dukes A" and "Dukes B" colon cancers are neoplasias that have invaded into the wall of the colon but have not spread into other tissues. Dukes A colon cancers are cancers that have not invaded beyond the submucosa. Dukes B colon cancers are subdivided into two groups: Dukes B1 and Dukes B2. "Dukes B1" colon cancers are neoplasias that have invaded up to but not through the muscularis propria. Dukes B2 colon cancers are cancers that have breached completely through the muscularis propria. Over a five year period, patients with Dukes A cancer who receive surgical treatment (i.e., removal of the affected tissue) have a greater than 90% survival rate. Over the same period, patients with Dukes B1 and Dukes B2 cancer receiving surgical treatment have a survival rate of about 85% and 75%, respectively. Dukes A, B1 and B2 cancers are also referred to as T1, T2 and T3-T4 cancers, respectively. "Dukes C" colon cancers are cancers that have spread to the regional lymph nodes, such as the lymph nodes of the gut. Patients with Dukes C cancer who receive surgical treatment alone have a 35% survival rate over a five year period, but this survival rate is increased to 60% in patients that receive chemotherapy. " Dukes D" colon cancers are cancers that have metastasized to other organs. The liver is the most common organ in which metastatic colon cancer is found. Patients with Dukes D colon cancer have a survival rate of less than 5% over a five year period, regardless of the treatment regimen. In general, colon neoplasia develops through one of at least three different pathways, termed chromosomal instability, microsatellite instability, and the CpG island methylator phenotype (CIMP). Although there is some overlap, these pathways tend to present somewhat different biological behavior. By understanding the pathway of tumor development, the target genes involved, and the mechanisms underlying the genetic instability, it is possible to implement strategies to detect and treat the different types of colon neoplasias.

In one aspect, this application is based at least in part, on the recognition that certain target genes may be silenced or inactivated by the differential methylation of CpG islands in the 5' flanking or promoter regions of the target gene. CpG islands are clusters of cytosine-guanosine residues in a DNA sequence, that are prominently represented in the 5-flanking region or promoter region of about half the genes in our genome. In particular, this application is based at least in part on the recognition that differential methylation of the SLC5A8 nucleotide

sequence may be indicative of a cancer (e.g., colon cancer, breast cancer, thyroid cancer, or stomach cancer).

As noted above, early detection of colon neoplasia, coupled with appropriate intervention, is important for increasing patient survival rates. Present systems for screening for colon neoplasia are deficient for a variety of reasons, including a lack of specificity and/or sensitivity (e.g., Fecal Occult Blood Test, flexible sigmoidoscopy) or a high cost and intensive use of medical resources (e.g., colonoscopy). Alternative systems for detection of colon neoplasia would be useful in a wide range of other clinical circumstances as well. For example, patients who receive surgical and/or pharmaceutical therapy for colon cancer may experience a relapse. It would be advantageous to have an alternative system for determining whether such patients have a recurrent or relapsed colon neoplasia. As a further example, an alternative diagnostic system would facilitate monitoring an increase, decrease or persistence of colon neoplasia in a patient known to have a colon neoplasia. A patient undergoing chemotherapy may be monitored to assess the effectiveness of the therapy.

In another aspect, the invention is also based, in part, on the discovery of a novel polynucleotide sequence encoding a novel sodium/solute symporter-like protein (SLC5A8). In particular, SLC5A8 is closely related to the human sodium iodide symporter (SLC5A5) and the human sodium-dependent multivitamin transporter (SLC5A6).

Cell surface receptors and transmembrane transporter systems facilitate communication between cells and their environment by direct exchange of chemicals between the intracellular and extracellular milieu. Distinct transporter systems (also called permeases, porters, transporters, carriers, and channel proteins) are specific for ions, small and medium size solutes and macromolecules. A major class of transporter proteins couple solute transport to the movement of other species (often cations, such as protons and sodium ions) either in the same direction (cotransporter or symporter) or in the opposite direction (counter transporter or antiporter). Sodium/solute symport is a widespread mechanism of solute transport across cytoplasmic membranes of prokaryotic and eukaryotic cells. Proteins that catalyze sodium/solute symport have been grouped into eleven families based on their degree of sequence similarities, their solute and cation specificities, size, topographical features, and evolutionary relationships (see, e.g., Reizer et al., (1994) Bichemica et Biphysica Acta, 1197:133-166). There are mixed families of transporters whose members differ in the choice of the coupling ion or catalyze symport or antiport processes.

Human sodium iodide transporter (NIS, or SLC5A5) is a best characterized member among the sodium/solute symporter superfamily. NIS localizes at the basolateral membrane and catalyses the active transport of iodide from blood into the cells using the inwardly directed sodium gradient with a 2 sodium 1 iodide stoichiometry. The tissue distribution of NIS includes the thyroid, salivary glands, stomach, thymus, and breast. Lower levels of expression of NIS are detected in the prostate, ovary, adrenal gland, lung, and heart. By contrast, the NIS gene has not been detected in the colon, orbital fibroblasts, or nasopharyngeal mucosa (see, e.g., Filetti et al., 1999, Eur J Endocrinol. 141:443–457). Abnormal NIS expression and/or iodide transport activity have been linked to many thyroid diseases including autoimmune thyroid diseases, thyroid nodular hyperplasia, thyroid adenoma, thyroid carcinoma, and congenital hypothyroidism, as well as non-thyroid diseases such as breast cancer and stomach cancer (Chung, 2002, J Nucl Med 43:1188-200).

Besides sequence homology to the human sodium iodide transporter, SLC5A8 transcript was found by Applicants to be expressed in the normal colon mucosa, kidney, lung, esophagus, small bowel, stomach, thyroid, and uterus. In addition, Applicants found that SLC5A8 may function as a sodium iodide transporter, and that differential methylation of SLC5A8 and/or reduced expression of SLC5A8 are linked to diseases such as colon cancer, breast cancer, and stomach cancer. Accordingly, the present invention relates to methods and compositions for detecting and treating such SLC5A8 associated cancers.

III. SLC5A8 polypeptides

In certain aspects, the invention provides a full-length SLC5A8 polypeptide (SEQ ID NO: 1) and functional variants thereof. Preferred functional variants of SLC5A8 polypeptides are those that have tumor suppressor activity or sodium transporter activity. In certain aspects, the present invention includes biologically-active fragments of the SLC5A8 protein and fusion proteins including at least a portion of the SLC5A8 protein. These include proteins with SLC5A8 activity that have amino acid substitutions or have sugars or other molecules attached to amino acid functional groups.

In certain embodiments, the present disclosure makes available isolated and/or purified forms of the SLC5A8 polypeptides, which are isolated from, or otherwise substantially free of, other proteins which might normally be associated with the protein or a particular complex including the protein. In certain embodiments, variant polypeptides have an amino acid

sequence that is at least 75% identical to an amino acid sequence as set forth in SEQ ID NO: 1. In other embodiments, the variant polypeptide has an amino acid sequence at least 80%, 85%, 90%, 95%, 97%, 98%, 99% or 100% identical to an amino acid sequence as set forth in SEQ ID NO: 1.

In certain aspects, variant SLC5A8 polypeptides are agonists or antagonists of the SLC5A8 polypeptide as set forth in SEQ ID NO: 1. Variants of these polypeptides may have a hyperactive or constitutive activity, or, act to prevent the tumor suppressor activity or sodium transporter activity of SLC5A8. For example, a truncated form lacking one or more domain may have a dominant negative effect.

In certain aspects, isolated peptidyl portions of the SLC5A8 polypeptide can be obtained by screening polypeptides recombinantly produced from the corresponding fragment of the nucleic acid encoding the polypeptide as set forth in SEQ ID NO: 1. In addition, fragments can be chemically synthesized using techniques known in the art such as conventional Merrifield solid phase f-Moc or t-Boc chemistry. The fragments can be produced (recombinantly or by chemical synthesis) and tested to identify those peptidyl fragments which can function as either agonists or antagonists of the SLC5A8 activity (e.g., tumor suppressor or sodium solute symporter).

The SLC5A8 protein is a transmembrane protein, with portions of the protein that are positioned outside the cell (the extracellular portions) and portions of the protein that are positioned inside the cell (the intracellular portions). Sequences and positions of the predicated thirteen transmembrane domains (TM1- TM13) are listed below.

TM1 (residues 10-32): FVVWDYVVFAGMLVISAAIGIYY (SEQ ID NO: 19)

TM2 (residues 52-74): MTAVPVALSLTASFMSAVTVLGT (SEQ ID NO: 20)

TM3 (residues 84-106): IFSIFAFTYFFVVVISAEVFLPV (SEQ ID NO: 21)

TM4 (residues 127-149): VRLCGTVLFIVQTILYTGIVIYA (SEQ ID NO: 22)

TM5 (residues 164-186): GAVVATGVVCTFYCTLGGLKAVI (SEQ ID NO: 23)

TM6 (residues 193-215): IGIMVAGFASVIIQAVVMQGGIS (SEQ ID NO: 24)

TM7 (residues 240-259): HTFWTIIIGGTFTWTSIYGV (SEQ ID NO: 25)

TM8 (residues 280-302): LYINLVGLWAILTCSVFCGLALY (SEQ ID NO: 26)

TM9 (residues 337-359): LPGLFVACAYSGTLSTVSSSINA (SEQ ID NO: 27)

TM10 (residues 380-402): SLSWISQGMSVVYGALCIGMAAL (SEQ ID NO: 28)

TM11 (residues 412-434): AALSVFGMVGGPLMGLFALGILV (SEQ ID NO: 29)

TM12 (residues 441-463): GALVGLMAGFAISLWVGIGAQIY (SEQ ID NO: 30)

TM13 (residues 519-541): LSYLYFSTVGTLVTLLVGILVSL (SEQ ID NO: 31)

Thus, certain embodiments of the invention include SLC5A8 fragments comprising a transmembrane domain as set forth in any of SEQ ID NOs: 19-21. In other embodiments, the present invention includes SLC5A8 fragments comprising an intracellular domain or an extracellular portion of the SLC5A8 protein.

In certain aspects, variant SLC5A8 polypeptides containing one or more fusion domains. Well known examples of such fusion domains include, for example, polyhistidine, Glu-Glu, glutathione S transferase (GST), thioredoxin, protein A, protein G, and an immunoglobulin heavy chain constant region (Fc), maltose binding protein (MBP), which are particularly useful for isolation of the fusion polypeptide by affinity chromatography. For the purpose of affinity purification, relevant matrices for affinity chromatography, such as glutathione-, amylase-, and nickel- or cobalt- conjugated resins are used. Many of such matrices are available in "kit" form, such as the Pharmacia GST purification system and the QIAexpress™ system (Qiagen) useful with (HIS_6) fusion partners. Another fusion domain well known in the art is green fluorescent protein (GFP). This fusion partner serves as a fluorescent "tag" which allows the fusion polypeptide of the invention to be identified by fluorescence microscopy or by flow cytometry. The GFP tag is useful when assessing subcellular localization of the fusion SLC5A8 polypeptide. The GFP tag is also useful for isolating cells which express the fusion SLC5A8 polypeptide by flow cytometric methods such as a fluorescence activated cell sorting (FACS). Fusion domains also include "epitope tags," which are usually short peptide sequences for which a specific antibody is available. Well known epitope tags for which specific monoclonal antibodies are readily available include FLAG, influenza virus haemagglutinin (HA), and c-myc tags. In some cases, the fusion domains have a protease cleavage site, such as for Factor Xa or Thrombin, which allow the relevant protease to partially digest the fusion SLC5A8 polypeptide

and thereby liberate the recombinant polypeptide therefrom. The liberated polypeptide can then be isolated from the fusion partner by subsequent chromatographic separation.

Different elements of fusion proteins may be arranged in any manner that is consistent with the desired functionality. For example, a SLC5A8 polypeptide may be placed C-terminal to a heterologous domain, or, alternatively, a heterologous domain may be placed C-terminal to a SLC5A8 polypeptide. The SLC5A8 and the heterologous domain need not be adjacent in a fusion protein, and additional domains or amino acid sequences may be included C- or N-terminal to either domain or between the domains.

It is also possible to modify the structure of the subject SLC5A8 polypeptides for such purposes as enhancing therapeutic or prophylactic efficacy, or stability (e.g., ex vivo shelf life and resistance to proteolytic degradation in vivo). Such modified polypeptides, when designed to retain at least one activity of the naturally occurring form of the protein, are considered functional equivalents of the SLC5A8 polypeptides described in more detail herein. Such modified polypeptides can be produced, for instance, by amino acid substitution, deletion or addition.

For instance, it is reasonable to expect, for example, that an isolated replacement of a leucine with an isoleucine or valine, an aspartate with a glutamate, a threonine with a serine, or a similar replacement of an amino acid with a structurally related amino acid (i.e., conservative mutations) will not have a major effect on the biological activity of the resulting molecule. Conservative replacements are those that take place within a family of amino acids that are related in their side chains (see, for example, Biochemistry, 2nd ed., Ed. by L. Stryer, W.H. Freeman and Co., 1981). Whether a change in the amino acid sequence of a polypeptide results in a functional homolog can be readily determined by assessing the ability of the variant polypeptide to produce a response in cells in a fashion similar to the wild-type protein. For instance, such variant forms of a SLC5A8 polypeptide can be assessed, e.g., for their ability to transport sodium solute or their ability to suppress tumor formation. Polypeptides in which more than one replacement has taken place can readily be tested in the same manner.

This invention further contemplates a method of generating sets of combinatorial mutants of the SLC5A8 polypeptides, as well as truncation mutants, and is especially useful for identifying potential variant sequences (e.g., homologs) that are functional in binding to a SLC5A8 polypeptide. The purpose of screening such combinatorial libraries may be to

generate, for example, SLC5A8 homologs which can act as either agonists or antagonist, or alternatively, which possess novel activities all together. Combinatorially-derived homologs can be generated which have a selective potency relative to a naturally occurring SLC5A8 polypeptide. Such proteins, when expressed from recombinant DNA constructs, can be used in gene therapy protocols. Likewise, mutagenesis can give rise to variants which have intracellular half-lives dramatically different than the corresponding wild-type protein. For example, the altered protein can be rendered either more stable or less stable to proteolytic degradation or other cellular process which result in destruction of, or otherwise inactivation of the SLC5A8 polypeptide of interest. Such variants, and the genes which encode them, can be utilized to alter SLC5A8 levels by modulating the half-life of the protein. For instance, a short half-life can give rise to more transient biological effects and, when part of an inducible expression system, can allow tighter control of recombinant SLC5A8 levels within the cell. As above, such proteins, and particularly their recombinant nucleic acid constructs, can be used in gene therapy protocols. In similar fashion, SLC5A8 homologs can be generated by the present combinatorial approach to act as antagonists, in that they are able to interfere with the ability of the corresponding wild-type protein to function.

In a representative embodiment of this method, the amino acid sequences for a population of SLC5A8 homologs are aligned, preferably to promote the highest homology possible. Such a population of variants can include, for example, homologs from one or more species, or homologs from the same species but which differ due to mutation. Amino acids which appear at each position of the aligned sequences may be selected to create a degenerate set of combinatorial sequences. In a preferred embodiment, the combinatorial library is produced by way of a degenerate library of genes encoding a library of polypeptides which each include at least a portion of potential SLC5A8 sequences. For instance, a mixture of synthetic oligonucleotides can be enzymatically ligated into gene sequences such that the degenerate set of potential SLC5A8 nucleotide sequences are expressible as individual polypeptides, or alternatively, as a set of larger fusion proteins (e.g., for phage display).

There are many ways by which the library of potential homologs can be generated from a degenerate oligonucleotide sequence. Chemical synthesis of a degenerate gene sequence can be carried out in an automatic DNA synthesizer, and the synthetic genes then be ligated into an appropriate gene for expression. The purpose of a degenerate set of genes is to provide, in one mixture, all of the sequences encoding the desired set of potential SLC5A8 sequences. The

f degenerate oligonucleotides is well known in the art (see for example, Narang, SA

(1983) Tetrahedron 39:3; Itakura et al., (1981) Recombinant DNA, Proc. 3rd Cleveland Sympos. Macromolecules, ed. AG Walton, Amsterdam: Elsevier pp273-289; Itakura et al., (1984) Annu. Rev. Biochem. 53:323; Itakura et al., (1984) Science 198:1056; Ike et al., (1983) Nucleic Acid Res. 11:477). Such techniques have been employed in the directed evolution of other proteins (see, for example, Scott et al., (1990) Science 249:386-390; Roberts et al., (1992) PNAS USA 89:2429-2433; Devlin et al., (1990) Science 249: 404-406; Cwirla et al., (1990) PNAS USA 87: 6378-6382; as well as U.S. Patent Nos: 5,223,409, 5,198,346, and 5,096,815).

Alternatively, other forms of mutagenesis can be utilized to generate a combinatorial library. For example, SLC5A8 variants (both agonist and antagonist forms) can be generated and isolated from a library by screening using, for example, alanine scanning mutagenesis and the like (Ruf et al., (1994) Biochemistry 33:1565-1572; Wang et al., (1994) J. Biol. Chem. 269:3095-3099; Balint et al., (1993) Gene 137:109-118; Grodberg et al., (1993) Eur. J. Biochem. 218:597-601; Nagashima et al., (1993) J. Biol. Chem. 268:2888-2892; Lowman et al., (1991) Biochemistry 30:10832-10838; and Cunningham et al., (1989) Science 244:1081-1085), by linker scanning mutagenesis (Gustin et al., (1993) Virology 193:653-660; Brown et al., (1992) Mol. Cell Biol. 12:2644-2652; McKnight et al., (1982) Science 232:316); by saturation mutagenesis (Meyers et al., (1986) Science 232:613); by PCR mutagenesis (Leung et al., (1989) Method Cell Mol Biol 1:11-19); or by random mutagenesis, including chemical mutagenesis, etc. (Miller et al., (1992) A Short Course in Bacterial Genetics, CSHL Press, Cold Spring Harbor, NY; and Greener et al., (1994) Strategies in Mol Biol 7:32-34). Linker scanning mutagenesis, particularly in a combinatorial setting, is an attractive method for identifying truncated (bioactive) forms of SLC5A8 polypeptides.

A wide range of techniques are known in the art for screening gene products of combinatorial libraries made by point mutations and truncations, and, for that matter, for screening cDNA libraries for gene products having a certain property. Such techniques will be generally adaptable for rapid screening of the gene libraries generated by the combinatorial mutagenesis of SLC5A8 variants. The most widely used techniques for screening large gene libraries typically comprises cloning the gene library into replicable expression vectors, transforming appropriate cells with the resulting library of vectors, and expressing the combinatorial genes under conditions in which detection of a desired activity facilitates relatively easy isolation of the vector encoding the gene whose product was detected. Each of the illustrative assays described below are amenable to high through-put analysis as necessary to

e numbers of degenerate sequences created by combinatorial mutagenesis techniques.

In an illustrative embodiment of a screening assay, candidate combinatorial gene products of one of the subject proteins are displayed on the surface of a cell or virus, and the ability of particular cells or viral particles to bind a SLC5A8 polypeptide is detected in a "panning assay." For instance, a library of SLC5A8 variants can be cloned into the gene for a surface membrane protein of a bacterial cell (Ladner et al.,, WO 88/06630; Fuchs et al., (1991) Bio/Technology 9:1370-1371; and Goward et al., (1992) TIBS 18:136-140), and the resulting fusion protein detected by panning, e.g., using a fluorescently labeled molecule which binds the SLC5A8 polypeptide, to score for potentially functional homologs. Cells can be visually inspected and separated under a fluorescence microscope, or, where the morphology of the cell permits, separated by a fluorescence-activated cell sorter.

In similar fashion, the gene library can be expressed as a fusion protein on the surface of a viral particle. For instance, in the filamentous phage system, foreign peptide sequences can be expressed on the surface of infectious phage, thereby conferring two significant benefits. First, since these phage can be applied to affinity matrices at very high concentrations, a large number of phage can be screened at one time. Second, since each infectious phage displays the combinatorial gene product on its surface, if a particular phage is recovered from an affinity matrix in low yield, the phage can be amplified by another round of infection. The group of almost identical E. coli filamentous phages M13, fd, and f1 are most often used in phage display libraries, as either of the phage gIII or gVIII coat proteins can be used to generate fusion proteins without disrupting the ultimate packaging of the viral particle (Ladner et al., PCT publication WO 90/02909; Garrard et al., PCT publication WO 92/09690; Marks et al., (1992) J. Biol. Chem. 267:16007-16010; Griffiths et al., (1993) EMBO J. 12:725-734; Clackson et al., (1991) Nature 352:624-628; and Barbas et al., (1992) PNAS USA 89:4457-4461).

In certain embodiments, the invention also provides for reduction of the subject SLC5A8 polypeptides to generate mimetics, e.g., peptide or non-peptide agents, which are able to mimic binding of the authentic protein to another cellular partner. Such mutagenic techniques as described above, as well as the thioredoxin system, are also particularly useful for mapping the determinants of a SLC5A8 polypeptide which participate in protein-protein interactions involved in, for example, binding of proteins involved in angiogenesis to each other. To illustrate, the critical residues of a SLC5A8 polypeptide which are involved in molecular recognition of a substrate protein can be determined and used to generate SLC5A8 polypeptide-derived peptidomimetics which bind to the substrate protein, and by inhibiting SLC5A8 binding, ›it its biological activity. By employing, for example, scanning mutagenesis to map

the amino acid residues of a SLC5A8 polypeptide which are involved in binding to another polypeptide, peptidomimetic compounds can be generated which mimic those residues involved in binding. For instance, non-hydrolyzable peptide analogs of such residues can be generated using benzodiazepine (e.g., see Freidinger et al., in Peptides: Chemistry and Biology, G.R. Marshall ed., ESCOM Publisher: Leiden, Netherlands, 1988), azepine (e.g., see Huffman et al., in Peptides: Chemistry and Biology, G.R. Marshall ed., ESCOM Publisher: Leiden, Netherlands, 1988), substituted gamma lactam rings (Garvey et al., in Peptides: Chemistry and Biology, G.R. Marshall ed., ESCOM Publisher: Leiden, Netherlands, 1988), keto-methylene pseudopeptides (Ewenson et al., (1986) J. Med. Chem. 29:295; and Ewenson et al., in Peptides: Structure and Function (Proceedings of the 9th American Peptide Symposium) Pierce Chemical Co. Rockland, IL, 1985), b-turn dipeptide cores (Nagai et al., (1985) Tetrahedron Lett 26:647; and Sato et al., (1986) J Chem Soc Perkin Trans 1:1231), and b-aminoalcohols (Gordon et al., (1985) Biochem Biophys Res Commun 126:419; and Dann et al., (1986) Biochem Biophys Res Commun 134:71).

In certain embodiments, the SLC5A8 polypeptides may further comprise post-translational or non-amino acid elements, such as hydrophobic modifications (e.g., polyethylene glycols or lipids), poly- or mono-saccharide modifications, phosphates, acetylations, etc. Effects of such elements on the functionality of a SLC5A8 polypeptide may be tested as described herein for other SLC5A8 variants.

In certain aspects, the present invention contemplates directly delivery of SLC5A8 polypeptides into a cell. Methods of directly introducing a polypeptide into a cell include, but are not limited to, protein transduction and protein therapy. For example, a protein transduction domain (PTD) can be fused to a nucleic acid encoding a SLC5A8 protein, and the fusion protein is expressed and purified. Fusion proteins containing the PTD are permeable to the cell membrane, and thus cells can be directly contacted with a fusion protein (Derossi et al. (1994) *Journal of Biological Chemistry* 269: 10444-10450; Han et al. (2000) *Molecules and Cells* 6: 728-732; Hall et al. (1996) *Current Biology* 6: 580-587; Theodore et al. (1995) *Journal of Neuroscience* 15: 7158-7167).

Although some protein transduction based methods rely on fusion of a polypeptide of interest to a sequence which mediates introduction of the protein into a cell, other protein transduction methods do not require covalent linkage of a protein of interest to a transduction domain. At least two commercially available reagents exist that mediate protein transduction

without covalent modification of the protein (Chariot™, produced by Active Motif, www.activemotif.com and Bioporter® Protein Delivery Reagent, produced by Gene Therapy Systems, www.genetherapysystems.com). Briefly, these protein transduction reagents can be used to deliver proteins, peptides and antibodies directly to cells including mammalian cells. Delivery of proteins directly to cells has a number of advantages. Firstly, many current techniques of gene delivery are based on delivery of a nucleic acid sequence which must be transcribed and/or translated by a cell before expression of the protein is achieved. This results in a time lag between delivery of the nucleic acid and expression of the protein. Direct delivery of a protein decreases this delay. Secondly, delivery of a protein often results in transient expression of the protein in a cell.

As outlined herein, protein transduction mediated by covalent attachment of a PTD to a protein can be used to deliver a protein to a cell. These methods require that individual proteins be covalently appended with PTD moieties. In contrast, methods such as Chariot™ and Bioporter® facilitate transduction by forming a noncovalent interaction between the reagent and the protein. Without being bound by theory, these reagents are thought to facilitate transit of the cell membrane, and following internalization into a cell the reagent and protein complex disassociates so that the protein is free to function in the cell.

IV. SLC5A8 nucleic acids

In certain aspects, the invention provides isolated and/or recombinant SLC5A8 nucleic acids encoding SLC5A8 polypeptides, for example, SEQ ID NOs: 3 and 4. The SLC5A8 polynucleotides may be single-stranded or double stranded. Such nucleic acids may be DNA or RNA molecules. The SLC5A8 nucleic acids are useful as diagnostic or therapeutic agents, such as for example, these nucleic acid molecules encode the SLC5A8 protein, and are useful in assaying for the presence of SLC5A8 transcripts in cancer cells (e.g., colon cancer cells, breast cancer cells, thyroid cancer cells, or stomach cancer cells).

SLC5A8 nucleic acids of the invention are further understood to include nucleic acids that comprise variants of SEQ ID NOs: 3 and 4. Variant nucleotide sequences include sequences that differ by one or more nucleotide substitutions, additions or deletions, such as allelic variants; and will, therefore, include coding sequences that differ from the nucleotide sequence of the coding sequence designated in SEQ ID NOs: 3 and 4. Optionally, a SLC5A8 nucleic acid of the invention will genetically complement a partial or complete SLC5A8 loss of

function phenotype. For example, a SLC5A8 nucleic acid of the invention may be expressed in a cell in which the endogenous SLC5A8 gene has been deleted, and the introduced SLC5A8 nucleic acid will mitigate a phenotype resulting from the gene deletion.

The present invention is based, at least in part, on the observation that SLC5A8 nucleotide sequences can be differentially methylated in certain SLC5A8-associated cancer, such as colon cancer, breast cancer, thyroid cancer or stomach cancer. Accordingly, certain aspects of the present invention provide SLC5A8 nucleic acids having certain regions that are differentially methylated in SLC5A8-associated cancer, for example, SEQ ID NOs: 12, 13, and 14, and fragments thereof. Detection of methylation in any one of such differentially methylated nucleic acid sequences would be indicative of a SLC5A8-associated cancer.

In certain embodiments, the application provides isolated or recombinant SLC5A8 nucleic acid sequences that are at least 80%, 85%, 90%, 95%, 97%, 98%, 99% or 100% identical to the SLC5A8 nucleic acid sequences (e.g., SEQ ID NOs: 3-4 and 12-14). One of ordinary skill in the art will appreciate that SLC5A8 nucleic acid sequences complementary to SEQ ID NOs: 3-4 and 12-14, and variants of SEQ ID NOs: 3-4 and 12-14 are also within the scope of this invention. In further embodiments, the SLC5A8 nucleic acid sequences of the invention can be isolated, recombinant, and/or fused with a heterologous nucleotide sequence, or in a DNA library.

In other embodiments, SLC5A8 nucleic acid sequences also include nucleotide sequences that hybridize under highly stringent conditions to the nucleotide sequences designated in SEQ ID NOs: 3-4 and 12-14, or fragments thereof. As discussed above, one of ordinary skill in the art will understand readily that appropriate stringency conditions which promote DNA hybridization can be varied. One of ordinary skill in the art will understand readily that appropriate stringency conditions which promote DNA hybridization can be varied. For example, one could perform the hybridization at 6.0 x sodium chloride/sodium citrate (SSC) at about 45 °C, followed by a wash of 2.0 x SSC at 50 °C. For example, the salt concentration in the wash step can be selected from a low stringency of about 2.0 x SSC at 50 °C to a high stringency of about 0.2 x SSC at 50 °C. In addition, the temperature in the wash step can be increased from low stringency conditions at room temperature, about 22 °C, to high stringency conditions at about 65 °C. Both temperature and salt may be varied, or temperature or salt concentration may be held constant while the other variable is changed. In one embodiment, the

invention provides nucleic acids which hybridize under low stringency conditions of 6 x SSC at room temperature followed by a wash at 2 x SSC at room temperature.

Isolated SLC5A8 nucleic acids which differ from the nucleic acids (e.g., SEQ ID NOs: 3-4 and 12-14) due to degeneracy in the genetic code are also within the scope of the invention. For example, a number of amino acids are designated by more than one triplet. Codons that specify the same amino acid, or synonyms (for example, CAU and CAC are synonyms for histidine) may result in "silent" mutations which do not affect the amino acid sequence of the protein. However, it is expected that DNA sequence polymorphisms that do lead to changes in the amino acid sequences of the subject proteins will exist among mammalian cells. One skilled in the art will appreciate that these variations in one or more nucleotides (up to about 3-5% of the nucleotides) of the nucleic acids encoding a particular protein may exist among individuals of a given species due to natural allelic variation. Any and all such nucleotide variations and resulting amino acid polymorphisms are within the scope of this invention.

In certain embodiments, the recombinant SLC5A8 nucleic acid may be operably linked to one or more regulatory nucleotide sequences in an expression construct. Regulatory nucleotide sequences will generally be appropriate for a host cell used for expression. Numerous types of appropriate expression vectors and suitable regulatory sequences are known in the art for a variety of host cells. Typically, said one or more regulatory nucleotide sequences may include, but are not limited to, promoter sequences, leader or signal sequences, ribosomal binding sites, transcriptional start and termination sequences, translational start and termination sequences, and enhancer or activator sequences. Constitutive or inducible promoters as known in the art are contemplated by the invention. The promoters may be either naturally occurring promoters, or hybrid promoters that combine elements of more than one promoter. An expression construct may be present in a cell on an episome, such as a plasmid, or the expression construct may be inserted in a chromosome. In a preferred embodiment, the expression vector contains a selectable marker gene to allow the selection of transformed host cells. Selectable marker genes are well known in the art and will vary with the host cell used.

In certain aspects, the application provides methylated forms of SLC5A8 nucleic acid sequences of SEQ ID NOs: 12-14 or fragments thereof, wherein the cytosine bases of the CpG islands present in said sequences are methylated. In other words, the SLC5A8 nucleic acid sequences of the present invention may be either in the methylated status (e.g., as seen in

SLC5A8-associated cancer tissues) or in the unmethylated status (e.g., as seen in normal tissues).

In certain embodiments, the present invention provides bisulfite-converted SLC5A8 template DNA sequences, for example, SEQ ID NOs: 15-18, and fragments thereof. Such bisulfite-converted SLC5A8 template DNA can be used for detecting the methylation status, for example, by an MSP reaction or by direct sequencing. These bisulfite-converted SLC5A8 sequences are also of use for designing primers for MS-PCR reactions that specifically detect methylated or unmethylated SLC5A8 templates following bisulfite conversion. In yet other embodiments, the bisulfite-converted SLC5A8 nucleotide sequences of the invention also include nucleotide sequences that will hybridize under highly stringent conditions to any nucleotide sequence selected from SEQ ID NOs: 15-18. In further aspects, the application provides methods for producing such bisulfite- converted nucleic acid sequences, for example, the application provides methods for treating a nucleotide sequence with a bisulfite agent such that the unmethylated cytosine bases are converted to a different nucleotide base such as a uracil.

The present invention also provides primers which can be used in PCR to obtain the SLC5A8 nucleic acids from cDNA The present invention also encompasses oligonucleotides that are useful as hybridization probes for detecting transcripts of the genes which encode the SLC5A8 protein Preferably, such oligonucleotides comprise at least 200 nucleotides. Such hybridization probes have a sequence which is at least 90% complementary with a contiguous sequence contained within the sense strand or antisense strand of a double stranded DNA molecule which encodes the SLC5A8 protein. Such hybridization probes bind to the sense strand or antisense under stringent conditions, preferably under highly stringent conditions. The probes are used in Northern assays to detect transcripts of SLC5A8 homologous genes and in Southern assays to detect SLC5A8 homologous genes. The identity of probes which are 200 nucleotides in length and have full complementarity with a portion of the sense or antisense strand of a double-stranded DNA molecule which encodes the SLC5A8 protein as set forth in SEQ ID NO: 1.

The various Sequence Identification Numbers that have been used in this application are summarized below in Table 1.

Table 1. Sequence Identification Numbers that have been used in this application.

SEQ ID NO	Description/ Name	Corresponding Figure
1	amino acid sequence of human SLC5A8 protein.	Figure 18.
2	genomic clone AC063951. Nucleotides 82200-83267 encompasses the promoter and/or exon 1 of the SLC5A8 gene, and referred to as the "SLC5A8 methylation target region."	Figure 1.
3	nucleotide sequence of the SLC5A8 mRNA transcript.	Figure 2.
4	nucleotide sequence of the SLC5A8 coding region.	Figure 23B.
5	3D41-Hpa2-190R	N/A.
6	3D41-Hpa2-633F	N/A.
7	3D41-Hpa2-82430F	N/A.
8	AS-unmeth-442s	N/A.
9	AS-unmeth-542as	N/A.
10	AS-meth-442-459s	N/A.
11	AS-meth-550as	N/A.
12	nucleotides 82200-83267 of AC063951, wild-type, sense strand.	Figure 4.
13	nucleotides 82200-83267 of AC063951, wild-type, antisense strand.	Figure 8.
14	nucleotides 300-600 of SEQ ID NO: 12, wild-type, antisense strand.	Figure 9.
15	nucleotides 82200-83267 of AC063951, antisense strand, bisulfite-converted/methylated.	Figure 10.
16	nucleotides 82200-83267 of AC063951, antisense strand, bisulfite-converted/unmethylated.	Figure 11.
17	nucleotides 82200-83267 of AC063951, sense strand, bisulfite-converted/methylated.	Figure 12.
18	nucleotides 82200-83267 of AC063951, sense strand, bisulfite-converted/unmethylated.	Figure 13.

V. SLC5A8 Expression Vectors

In certain aspects, nucleic acids encoding SLC5A8 polypeptides and variants thereof may be used to increase SLC5A8 expression in an organism or cell by direct delivery of the nucleic acid. A nucleic acid therapy construct of the present invention can be delivered, for example, as an expression plasmid which, when transcribed in the cell, produces RNA which encodes a SLC5A8 polypeptide.

In another aspect of the invention, the subject nucleic acid is provided in an expression vector comprising a nucleotide sequence encoding a subject SLC5A8 polypeptide and operably linked to at least one regulatory sequence. Regulatory sequences are art-recognized and are selected to direct expression of the SLC5A8 polypeptide. Accordingly, the term regulatory sequence includes promoters, enhancers, and other expression control elements. Exemplary regulatory sequences are described in Goeddel; *Gene Expression Technology*: *Methods in Enzymology*, Academic Press, San Diego, CA (1990). For instance, any of a wide variety of expression control sequences that control the expression of a DNA sequence when operatively linked to it may be used in these vectors to express DNA sequences encoding a SLC5A8 polypeptide. Such useful expression control sequences, include, for example, the early and late promoters of SV40, tet promoter, adenovirus or cytomegalovirus immediate early promoter, the lac system, the trp system, the TAC or TRC system, T7 promoter whose expression is directed by T7 RNA polymerase, the major operator and promoter regions of phage lambda , the control regions for fd coat protein, the promoter for 3-phosphoglycerate kinase or other glycolytic enzymes, the promoters of acid phosphatase, e.g., Pho5, the promoters of the yeast α-mating factors, the polyhedron promoter of the baculovirus system and other sequences known to control the expression of genes of prokaryotic or eukaryotic cells or their viruses, and various combinations thereof. It should be understood that the design of the expression vector may depend on such factors as the choice of the host cell to be transformed and/or the type of protein desired to be expressed. Moreover, the vector's copy number, the ability to control that copy number and the expression of any other protein encoded by the vector, such as antibiotic markers, should also be considered.

As will be apparent, the subject gene constructs can be used to cause expression of the subject SLC5A8 polypeptides in cells propagated in culture, e.g., to produce proteins or polypeptides, including fusion proteins or polypeptides, for purification.

This invention also pertains to a host cell transfected with a recombinant gene including a coding sequence for one or more of the subject SLC5A8 polypeptides. The host cell may be any prokaryotic or eukaryotic cell. For example, a polypeptide of the present invention may be expressed in bacterial cells such as *E. coli*, insect cells (e.g., using a baculovirus expression system), yeast, or mammalian cells. Other suitable host cells are known to those skilled in the art.

Accordingly, the present invention further pertains to methods of producing the subject SLC5A8 polypeptides. For example, a host cell transfected with an expression vector encoding a SLC5A8 polypeptide can be cultured under appropriate conditions to allow expression of the polypeptide to occur. The polypeptide may be secreted and isolated from a mixture of cells and medium containing the polypeptide. Alternatively, the polypeptide may be retained cytoplasmically or in a membrane fraction and the cells harvested, lysed and the protein isolated. A cell culture includes host cells, media and other byproducts. Suitable media for cell culture are well known in the art. The polypeptide can be isolated from cell culture medium, host cells, or both using techniques known in the art for purifying proteins, including ion-exchange chromatography, gel filtration chromatography, ultrafiltration, electrophoresis, and immunoaffinity purification with antibodies specific for particular epitopes of the polypeptide. In a preferred embodiment, the SLC5A8 polypeptide is a fusion protein containing a domain which facilitates its purification, such as a SLC5A8-GST fusion protein, SLC5A8-intein fusion protein, SLC5A8-cellulose binding domain fusion protein, SLC5A8-polyhistidine fusion protein, etc.

A recombinant SLC5A8 nucleic acid can be produced by ligating the cloned gene, or a portion thereof, into a vector suitable for expression in either prokaryotic cells, eukaryotic cells (yeast, avian, insect or mammalian), or both. Expression vehicles for production of a recombinant SLC5A8 polypeptides include plasmids and other vectors. For instance, suitable vectors for the expression of a SLC5A8 polypeptide include plasmids of the types: pBR322-derived plasmids, pEMBL-derived plasmids, pEX-derived plasmids, pBTac-derived plasmids and pUC-derived plasmids for expression in prokaryotic cells, such as *E. coli*.

The preferred mammalian expression vectors contain both prokaryotic sequences to facilitate the propagation of the vector in bacteria, and one or more eukaryotic transcription units that are expressed in eukaryotic cells. The pcDNAI/amp, pcDNAI/neo, pRc/CMV, pSV2gpt, pSV2neo, pSV2-dhfr, pTk2, pRSVneo, pMSG, pSVT7, pko-neo and pHyg derived vectors are

examples of mammalian expression vectors suitable for transfection of eukaryotic cells. Some of these vectors are modified with sequences from bacterial plasmids, such as pBR322, to facilitate replication and drug resistance selection in both prokaryotic and eukaryotic cells. Alternatively, derivatives of viruses such as the bovine papilloma virus (BPV-1), or Epstein-Barr virus (pHEBo, pREP-derived and p205) can be used for transient expression of proteins in eukaryotic cells. Examples of other viral (including retroviral) expression systems can be found below in the description of gene therapy delivery systems. The various methods employed in the preparation of the plasmids and transformation of host organisms are well known in the art. For other suitable expression systems for both prokaryotic and eukaryotic cells, as well as general recombinant procedures, see *Molecular Cloning A Laboratory Manual*, 2nd Ed., ed. by Sambrook, Fritsch and Maniatis (Cold Spring Harbor Laboratory Press, 1989) Chapters 16 and 17. In some instances, it may be desirable to express the recombinant SLC5A8 polypeptide by the use of a baculovirus expression system. Examples of such baculovirus expression systems include pVL-derived vectors (such as pVL1392, pVL1393 and pVL941), pAcUW-derived vectors (such as pAcUW1), and pBlueBac-derived vectors (such as the ß-gal containing pBlueBac III).

In another embodiment, a fusion gene coding for a purification leader sequence, such as a poly-(His)/enterokinase cleavage site sequence at the N-terminus of the desired portion of the recombinant SLC5A8 protein, can allow purification of the expressed fusion protein by affinity chromatography using a Ni^{2+} metal resin. The purification leader sequence can then be subsequently removed by treatment with enterokinase to provide the purified SLC5A8 polypeptide (e.g., see Hochuli et al., (1987) *J. Chromatography* 411:177; and Janknecht et al., *PNAS USA* 88:8972).

Techniques for making fusion genes are well known. Essentially, the joining of various DNA fragments coding for different polypeptide sequences is performed in accordance with conventional techniques, employing blunt-ended or stagger-ended termini for ligation, restriction enzyme digestion to provide for appropriate termini, filling-in of cohesive ends as appropriate, alkaline phosphatase treatment to avoid undesirable joining, and enzymatic ligation. In another embodiment, the fusion gene can be synthesized by conventional techniques including automated DNA synthesizers. Alternatively, PCR amplification of gene fragments can be carried out using anchor primers which give rise to complementary overhangs between two consecutive gene fragments which can subsequently be annealed to generate a chimeric

gene sequence (see, for example, *Current Protocols in Molecular Biology*, eds. Ausubel et al., John Wiley & Sons: 1992).

VI. Antibodies

Another aspect of the invention pertains to an antibody reactive with a SLC5A8 polypeptide, preferably antibodies that are specifically reactive with SLC5A8 polypeptide. For example, by using immunogens derived from a SLC5A8 polypeptide, anti-protein/anti-peptide antisera or monoclonal antibodies can be made by standard protocols (see, for example, Antibodies: A Laboratory Manual ed. by Harlow and Lane (Cold Spring Harbor Press: 1988)). A mammal, such as a mouse, a hamster or rabbit can be immunized with an immunogenic form of the peptide (e.g., a SLC5A8 polypeptide or an antigenic fragment which is capable of eliciting an antibody response, or a fusion protein). Techniques for conferring immunogenicity on a protein or peptide include conjugation to carriers or other techniques well known in the art. An immunogenic portion of a SLC5A8 polypeptide can be administered in the presence of adjuvant. The progress of immunization can be monitored by detection of antibody titers in plasma or serum. Standard ELISA or other immunoassays can be used with the immunogen as antigen to assess the levels of antibodies. In a preferred embodiment, the subject antibodies are immunospecific for antigenic determinants of a SLC5A8 polypeptide as set forth in SEQ ID NO: 1.

In one embodiment, antibodies are specific for the SLC5A8 protein as encoded by nucleic acid sequences as set forth in SEQ ID NOs: 3 and 4. In other embodiments, an antibody is immunoreactive with one or more proteins having an amino acid sequence that is at least 85%, 90%, 95%, 98%, 99%, 99.3%, 99.5%, 99.7% or 100% identical to an amino acid sequence as set forth in SEQ ID NO: 1.

In another embodiment, antibodies of the invention are specific for the extracellular portion of the SLC5A8 protein. In a set of exemplary embodiments, an antibody binds to an extracellular portion of SEQ ID NO: 1. In another embodiment, antibodies of the invention are specific for the intracellular portion or the transmembrane portion of the SLC5A8 protein. In a further embodiment, antibodies of the invention are specific for the soluble SLC5A8 protein and variants thereof.

Following immunization of an animal with an antigenic preparation of a SLC5A8 polypeptide, anti-SLC5A8 antisera can be obtained and, if desired, polyclonal anti-SLC5A8

antibodies can be isolated from the serum. To produce monoclonal antibodies, antibody-producing cells (lymphocytes) can be harvested from an immunized animal and fused by standard somatic cell fusion procedures with immortalizing cells such as myeloma cells to yield hybridoma cells. Such techniques are well known in the art, and include, for example, the hybridoma technique (originally developed by Kohler and Milstein, (1975) Nature, 256: 495-497), the human B cell hybridoma technique (Kozbar et al., (1983) Immunology Today, 4: 72), and the EBV-hybridoma technique to produce human monoclonal antibodies (Cole et al., (1985) Monoclonal Antibodies and Cancer Therapy, Alan R. Liss, Inc. pp. 77-96). Hybridoma cells can be screened immunochemically for production of antibodies specifically reactive with a SLC5A8 polypeptide of the present invention and monoclonal antibodies isolated from a culture comprising such hybridoma cells. In one embodiment, anti-SLC5A8 antibodies specifically react with the protein encoded by a nucleic acid having the sequence of SEQ ID NO: 3 or 4.

The term "antibody" as used herein is intended to include fragments thereof which are also specifically reactive with a subject SLC5A8 polypeptide. Antibodies can be fragmented using conventional techniques and the fragments screened for utility in the same manner as described above for whole antibodies. For example, $F(ab)_2$ fragments can be generated by treating antibody with pepsin. The resulting $F(ab)_2$ fragment can be treated to reduce disulfide bridges to produce Fab fragments. The antibody of the present invention is further intended to include bispecific, single-chain, and chimeric and humanized molecules having affinity for a SLC5A8 polypeptide conferred by at least one CDR region of the antibody. In preferred embodiments, the antibody further comprises a label attached thereto and able to be detected (e.g., the label can be a radioisotope, fluorescent compound, enzyme or enzyme co-factor).

In certain preferred embodiments, an antibody of the invention is a monoclonal antibody, and in certain embodiments, the invention makes available methods for generating novel antibodies. For example, a method for generating a monoclonal antibody that binds specifically to a SLC5A8 polypeptide may comprise administering to a mouse an amount of an immunogenic composition comprising the SLC5A8 polypeptide effective to stimulate a detectable immune response, obtaining antibody-producing cells (e.g., cells from the spleen) from the mouse and fusing the antibody–producing cells with myeloma cells to obtain antibody-producing hybridomas, and testing the antibody-producing hybridomas to identify a hybridoma that produces a monocolonal antibody that binds specifically to the SLC5A8 polypeptide. Once obtained, a hybridoma can be propagated in a cell culture, optionally in culture conditions where

the hybridoma-derived cells produce the monoclonal antibody that binds specifically to the SLC5A8 polypeptide. The monoclonal antibody may be purified from the cell culture.

Anti-SLC5A8 antibodies can be used, e.g., to detect SLC5A8 polypeptides in biological samples and/or to monitor SLC5A8 polypeptide levels in an individual. The level of SLC5A8 polypeptide maybe measured in a variety of sample types such as, for example, in cells , stools, and/or in bodily fluid, such as in whole blood samples, blood serum, blood plasma and urine. The adjective "specifically reactive with" as used in reference to an antibody is intended to mean, as is generally understood in the art, that the antibody is sufficiently selective between the antigen of interest (e.g., a SLC5A8 polypeptide) and other antigens that are not of interest that the antibody is useful for, at minimum, detecting the presence of the antigen of interest in a particular type of biological sample. In certain methods employing the antibody, a higher degree of specificity in binding may be desirable. For example, an antibody for use in detecting a low abundance protein of interest in the presence of one or more very high abundance protein that are not of interest may perform better if it has a higher degree of selectivity between the antigen of interest and other cross-reactants. Monoclonal antibodies generally have a greater tendency (as compared to polyclonal antibodies) to discriminate effectively between the desired antigens and cross-reacting polypeptides. In addition, an antibody that is effective at selectively identifying an antigen of interest in one type of biological sample (e.g., a stool sample) may not be as effective for selectively identifying the same antigen in a different type of biological sample (e.g., a blood sample). Likewise, an antibody that is effective at identifying an antigen of interest in a purified protein preparation that is devoid of other biological contaminants may not be as effective at identifying an antigen of interest in a crude biological sample, such as a blood or urine sample. Accordingly, in preferred embodiments, the application provides antibodies that have demonstrated specificity for a SLC5A8 protein in a sample type that is likely to be the sample type of choice for use of the antibody. In a particularly preferred embodiment, the application provides antibodies that bind specifically to a SLC5A8 polypeptide in a protein preparation from blood (optionally serum or plasma) from a patient that has a SLC5A8 associated cancer or that bind specifically in a crude blood sample (optionally a crude serum or plasma sample).

One characteristic that influences the specificity of an antibody:antigen interaction is the affinity of the antibody for the antigen. Although the desired specificity may be reached with a range of different affinities, generally preferred antibodies will have an affinity (a dissociation f about 10^{-6}, 10^{-7}, 10^{-8}, 10^{-9} or less.

In addition, the techniques used to screen antibodies in order to identify a desirable antibody may influence the properties of the antibody obtained. For example, an antibody to be used for certain therapeutic purposes will preferably be able to target a particular cell type. Accordingly, to obtain antibodies of this type, it may be desirable to screen for antibodies that bind to cells that express the antigen of interest (e.g., by fluorescence activated cell sorting). Likewise, if an antibody is to be used for binding an antigen in solution, it may be desirable to test solution binding. A variety of different techniques are available for testing interaction between antibodies and antigens to identify particularly desirable antibodies. Such techniques include ELISAs, surface plasmon resonance binding assays (e.g., the Biacore binding assay, Bia-core AB, Uppsala, Sweden), sandwich assays (e.g., the paramagnetic bead system of IGEN International, Inc., Gaithersburg, Maryland), western blots, immunoprecipitation assays, and immunohistochemistry.

In certain embodiment, antibodies of the invention may be useful as diagnostic or therapeutic agents for detecting or treating SLC5A8-associated diseases (e.g., cancers). The diagnostic method comprises the steps of contacting a sample of test cells or a protein extract thereof with immunospecific anti-SLC5A8 antibodies and assaying for the formation of a complex between the antibodies and a protein in the sample. Formation of low levels of complex in the test cell as compared to the normal cells indicates that the test cell is cancerous.

VII. Transgenic Animals

Another aspect of the invention features transgenic non-human animals which express a heterologous SLC5A8 gene, e.g., having a sequence of SEQ ID NO: 3 or 4, or fragments thereof. In another aspect, the invention features transgenic non-human animals which have had one or both copies of the endogenous SLC5A8 genes disrupted in at least one of the tissue or cell-types of the animal. In one embodiment, the transgenic non-human animals is a mammal such as a mouse, rat, rabbit, goat, sheep, dog, cat, cow or non-human primate. Without being bound to theory, it is proposed that such an animal may display a phenomenon associated with reduced or increased chance of cancer development (e.g., colon cancer, breast cancer, thyroid cancer, or stomach cancer). Accordingly, such a transgenic animal may serve as a useful animal model to study the progression of cancer diseases.

The term "transgene" is used herein to describe genetic material that has been or is about to be artificially inserted into the genome of a mammalian cell, particularly a mammalian cell of

a living animal. The transgene is used to transform a cell, meaning that a permanent or transient genetic change, preferably a permanent genetic change, is induced in a cell following incorporation of exogenous DNA. A permanent genetic change is generally achieved by introduction of the DNA into the genome of the cell. Vectors for stable integration include plasmids, retroviruses and other animal viruses, YACs, and the like. Of interest are transgenic mammals, e.g., cows, pigs, goats, horses, etc., and particularly rodents, e.g., rats, mice, etc. Preferably, the transgenic-animals are mice.

Transgenic animals comprise an exogenous nucleic acid sequence present as an extrachromosomal element or stably integrated in all or a portion of its cells, especially in germ cells. Unless otherwise indicated, it will be assumed that a transgenic animal comprises stable changes to the germline sequence. During the initial construction of the animal, "chimeras" or "chimeric animals" are generated, in which only a subset of cells have the altered genome. Chimeras are primarily used for breeding purposes in order to generate the desired transgenic animal. Animals having a heterozygous alteration are generated by breeding of chimeras. Male and female heterozygotes are typically bred to generate homozygous animals.

The exogenous gene is usually either from a different species than the animal host, or is otherwise altered in its coding or non-coding sequence. The introduced gene may be a wild-type gene, naturally occurring polymorphism, or a genetically manipulated sequence, for example having deletions, substitutions or insertions in the coding or non-coding regions. Where the introduced gene is a coding sequence, it is usually operably linked to a promoter, which may be constitutive or inducible, and other regulatory sequences required for expression in the host animal.

In one aspect of the invention, a SLC5A8 transgene can encode the wild-type form of the protein, homologs thereof, as well as antisense constructs. A SLC5A8 transgene can also encode a soluble form of SLC5A8 that has tumor suppressor activity or sodium solute transporter activity.

It may be desirable to express the heterologous SLC5A8 transgene conditionally such that either the timing or the level of SLC5A8 gene expression can be regulated. Such conditional expression can be provided using prokaryotic promoter sequences which require prokaryotic proteins to be simultaneous expressed in order to facilitate expression of the

SLC5A8 transgene. Exemplary promoters and the corresponding trans-activating prokaryotic proteins are given in U.S. Patent No. 4,833,080.

Moreover, transgenic animals exhibiting tissue specific expression can be generated, for example, by inserting a tissue specific regulatory element, such as an enhancer, into the transgene. For example, the endogenous SLC5A8 gene promoter or a portion thereof can be replaced with another promoter and/or enhancer, e.g., a CMV or a Moloney murine leukemia virus (MLV) promoter and/or enhancer.

Transgenic animals containing an inducible SLC5A8 transgene can be generated using inducible regulatory elements (e.g., metallothionein promoter), which are well-known in the art. SLC5A8 transgene expression can then be initiated in these animals by administering to the animal a compound which induces gene expression (e.g., heavy metals). Another preferred inducible system comprises a tetracycline-inducible transcriptional activator (U.S. Patent Nos. 5,654,168 and 5,650,298).

The present invention provides transgenic animals that carry the transgene in all their cells, as well as animals that carry the transgene in some, but not all cells, i.e., mosaic animals. The transgene can be integrated as a single transgene or in tandem, e.g., head to head tandems, or head to tail or tail to tail or as multiple copies.

The successful expression of the transgene can be detected by any of several means well known to those skilled in the art. Non-limiting examples include Northern blot, in situ hybridization of mRNA analysis, Western blot analysis, immunohistochemistry, and FACS analysis of protein expression.

In a further aspect, the invention features non-human animal cells containing a SLC5A8 transgene, preferentially a human SLC5A8 transgene. For example, the animal cell (e.g., somatic cell or germ cell (i.e., egg or sperm)) can be obtained from the transgenic animal. Transgenic somatic cells or cell lines can be used, for example, in drug screening assays. Transgenic germ cells, on the other hand, can be used in generating transgenic progeny.

Although not necessary to the operability of the invention, the transgenic animals described herein may comprise alterations to endogenous genes in addition to, or alternatively, to the genetic alterations described above. For example, the host animals may be either "knockouts" or "knockins" for the SLC5A8 gene. Knockouts have a partial or complete loss of

function in one or both alleles of an endogenous gene of interest. Knockins have an introduced transgene with altered genetic sequence and/or function from the endogenous gene. The two may be combined, for example, such that the naturally occurring gene is disabled, and an altered form introduced. For example, it may be desirable to knockout the host animal's endogenous SLC5A8 gene, while introducing an exogenous SLC5A8 gene (e.g., a human SLC5A8 gene).

In a knockout, preferably the target gene expression is undetectable or insignificant. For example, a knock-out of a SLC5A8 gene means that function of the SLC5A8 has been substantially decreased so that expression is not detectable or only present at insignificant levels. This may be achieved by a variety of mechanisms, including introduction of a disruption of the coding sequence, e.g., insertion of one or more stop codons, insertion of a DNA fragment, deletion of coding sequence, substitution of stop codons for coding sequence, etc. In some cases, the exogenous transgene sequences are ultimately deleted from the genome, leaving a net change to the native sequence. Different approaches may be used to achieve the "knock-out." A chromosomal deletion of all or part of the native gene may be induced, including deletions of the non-coding regions, particularly the promoter region, 3' regulatory sequences, enhancers, or deletions of gene that activate expression of APP genes. A functional knock-out may also be achieved by the introduction of an anti-sense construct that blocks expression of the native genes (for example, see Li and Cohen (1996) Cell 85:319-329). "Knock-outs" also include conditional knock-outs, for example, where alteration of the target gene occurs upon exposure of the animal to a substance that promotes target gene alteration, introduction of an enzyme that promotes recombination at the target gene site (e.g., Cre in the Cre-lox system), or other method for directing the target gene alteration postnatally.

A "knockin" of a target gene means an alteration in a host cell genome that results in altered expression or function of a native target gene. Increased (including ectopic) or decreased expression may be achieved by introduction of an additional copy of the target gene, or by operatively inserting a regulatory sequence that provides for enhanced expression of an endogenous copy of the target gene. These changes may be constitutive or conditional, i.e., dependent on the presence of an activator or repressor. The use of knockin technology may be combined with production of exogenous sequences to produce the transgenic animals of the invention.

DNA constructs for random integration need not include regions of homology to mediate recombination. Where homologous recombination is desired, the DNA constructs will comprise

at least a portion of the target gene with the desired genetic modification, and will include regions of homology to the target locus. Conveniently, markers for positive and negative selection are included. Methods for generating cells having targeted gene modifications through homologous recombination are known in the art. For various techniques for transfecting mammalian cells, see Keown et al. (1990) Methods in Enzymology 185:527-537.

For embryonic stem (ES) cells, an ES cell line may be employed, or embryonic cells may be obtained freshly from a host, e.g., mouse, rat, or guinea pig. Such cells are grown on an appropriate fibroblast-feeder layer or grown in the presence of appropriate growth factors, such as leukemia inhibiting factor (LIF). When ES cells have been transformed, they may be used to produce transgenic animals. After transformation, the cells are plated onto a feeder layer in an appropriate medium. Cells containing the construct may be detected by employing a selective medium. After sufficient time for colonies to grow, they are picked and analyzed for the occurrence of homologous recombination or integration of the construct. Those colonies that are positive may then be used for embryo manipulation and blastocyst injection. Blastocysts are obtained from 4 to 6 week old superovulated females. The ES cells are trypsinized, and the modified cells are injected into the blastocoel of the blastocyst. After injection, the blastocysts are returned to each uterine horn of pseudopregnant females. Females are then allowed to go to term and the resulting litters screened for mutant cells having the construct. By providing for a different phenotype of the blastocyst and the ES cells, chimeric progeny can be readily detected.

The chimeric animals are screened for the presence of the modified gene and males and females having the modification are mated to produce homozygous progeny. If the gene alterations cause lethality at some point in development, tissues or organs can be maintained as allogeneic or congenic grafts or transplants, or in in vitro culture.

The transgenic animals of the present invention may be an animal model for a SLC5A8-associated disease (e.g., cancer), and display cancer-related phenotypes (e.g., colon cancer, breast cancer, thyroid cancer, or stomach cancer), depending on different alleles generated. Accordingly, such transgenic animals can be used in in vivo assays to identify cancer therapeutics. In an exemplary embodiment, the assay comprises administering a test compound to a transgenic animal of the invention, and comparing a phenotypic change in cancer development in the animal relative to a transgenic animal which has not received the test compound.

To illustrate, the transgenic animals and cell lines are particularly useful in screening compounds that have potential as prophylactic or therapeutic treatments of diseases such as may involve aberrant expression, or loss, of the SLC5A8 gene. Screening for a useful drug would involve administering the candidate drug over a range of doses to the transgenic animal, and assaying at various time points for the effect(s) of the drug on the disease or disorder being evaluated. Alternatively, or additionally, the drug could be administered prior to or simultaneously with exposure to induction of the disease, if applicable.

In one embodiment, candidate compounds are screened by being administered to the transgenic animal, over a range of doses, and evaluating the animal's physiological response to the compound(s) over time. Administration may be oral, or by suitable injection, depending on the chemical nature of the compound being evaluated. In some cases, it may be appropriate to administer the compound in conjunction with co-factors that would enhance the efficacy of the compound.

In screening cell lines derived from the subject transgenic animals for compounds useful in treating various disorders, the test compound is added to the cell culture medium at the appropriate time, and the cellular response to the compound is evaluated over time using the appropriate biochemical and/or histological assays. In some cases, it may be appropriate to apply the compound of interest to the culture medium in conjunction with co-factors that would enhance the efficacy of the compound.

In another aspect, the animals of this invention can be used as a source of cells, differentiated or precursor, which can be immortalized in cell culture. Cells in which the normal function of the SLC5A8 protein is altered by a transgene may be isolated from potentially any tissue of the animal, as well as form animals at any developmental stage, e.g. embryonic to adult. The subject transgenic animals can, accordingly, be used as a source of material for the growth, identification, purification and detailed analysis of, inter alia, precursor cells, including stem cells and pluripotent progenitor cells for a variety of tissues.

Vectors used for transforming animal embryos are constructed using methods well known in the art, including, without limitation, the standard techniques of restriction endonuclease digestion, ligation, plasmid and DNA and RNA purification, DNA sequencing, and the like as described, for example in Sambrook, Fritsch, and Maniatis, eds., Molecular Cloning: A Laboratory Manual., (Cold Spring Harbor Laboratory Press, Cold Spring Harbor,

N.Y. ›1989!). Most practitioners are familiar with the standard resource materials as well as specific conditions and procedures.

VIII. Screening Assays

The invention provides methods (also referred to herein as "screening assays") for identifying modulators, i.e., candidate or test compounds or agents (e.g., proteins, peptides, peptidomimetics, peptoids, small molecules or other drugs) which bind to SLC5A8 proteins, have a stimulatory or inhibitory effect on, for example, SLC5A8 expression or SLC5A8 activity, or have a stimulatory or inhibitory effect on, for example, the expression or activity of a SLC5A8 substrate. Compounds thus identified can be used to modulate the activity of target gene products (e.g., the SLC5A8 gene) in a therapeutic protocol, to elaborate the biological function of the target gene product, or to identify compounds that disrupt normal target gene interactions. Given that the SLC5A8 polypeptide is a transmembrane protein, agents that bind to a SLC5A8 polypeptide may include its natural ligands, downstream signaling molecules, and other endogenous polypeptides as well as artificial compounds. In one embodiment, an assay detects agents which inhibit interaction of the subject SLC5A8 polypeptides with a SLC5A8-associated protein. A wide variety of assays may be used for this purpose, including labeled in vitro protein-protein binding assays, interaction trap assay, immunoassays for protein binding, and the like.

Given the role of SLC5A8 in transporting sodium solute and in cancer development, the agents that bind to SLC5A8 as well as the agents that interfere with SLC5A8 binding to SLC5A8-associated proteins may be able to modulate transporting sodium solute or cancer development. Accordingly, one aspect of the invention provides a method for assessing the ability of an agent to modulate transporting sodium solute or cancer development, comprising: 1) combining: a first polypeptide including at least a portion of a SLC5A8 polypeptide, a second polypeptide including at least a portion of a SLC5A8-associated protein that interacts with the first polypeptide, and an agent, under conditions wherein the first polypeptide interacts with the second polypeptide in the absence of said agent, 2) determining if said agent interferes with the interaction, and 3) for an agent that interferes with the interaction, further assessing its ability to interfere with SLC5A8's ability to transport sodium solute or suppress tumor development.

In one embodiment, an activity (e.g., the sodium solute transporting activity) of a SLC5A8 protein can be assayed as follows. Xenopus laevis oocytes are injected with mRNA

encoding the SLC5A8 protein or a eukaryotic expression vector able to express such an mRNA, using a Drummond Nanoject (Drummond Scientific, Broomall, Pa. into the animal pole of defolliculated oocytes as described by Swick et al. ((1992) Proc. Natl. Acad. Sci. USA. 89:1812-1816). The injected oocytes are then transferred to microtiter wells about 12 to 24 hours prior to being assayed. The transporter function of oocyte-expressed SLC5A8 polypeptide is assessed by sodium uptakes as described (see e.g., Romera et al. (2000) J. Biol. Chem. 275:24552-24559; Sciortino et al. (1999) Am. J. Physiol. 277:F611-623).

A variety of assay formats will suffice and, in light of the present disclosure, those not expressly described herein will nevertheless be comprehended by one of ordinary skill in the art. Assay formats which approximate such conditions as formation of protein complexes, enzymatic activity, may be generated in many different forms, and include assays based on cell-free systems, e.g., purified proteins or cell lysates, as well as cell-based assays which utilize intact cells. Simple binding assays can also be used to detect agents which bind to SLC5A8. Such binding assays may also identify agents that act by disrupting the interaction between a SLC5A8 polypeptide and a SLC5A8 interacting protein. Agents to be tested can be produced, for example, by bacteria, yeast or other organisms (e.g., natural products), produced chemically (e.g., small molecules, including peptidomimetics), or produced recombinantly. In a preferred embodiment, the test agent is a small organic molecule, e.g., other than a peptide or oligonucleotide, having a molecular weight of less than about 2,000 daltons.

In many drug screening programs which test libraries of compounds and natural extracts, high throughput assays are desirable in order to maximize the number of compounds surveyed in a given period of time. Assays of the present invention which are performed in cell-free systems, such as may be developed with purified or semi-purified proteins or with lysates, are often preferred as "primary" screens in that they can be generated to permit rapid development and relatively easy detection of an alteration in a molecular target which is mediated by a test compound. Moreover, the effects of cellular toxicity and/or bioavailability of the test compound can be generally ignored in the in vitro system, the assay instead being focused primarily on the effect of the drug on the molecular target as may be manifest in an alteration of binding affinity with other proteins or changes in enzymatic properties of the molecular target.

In preferred in vitro embodiments of the present assay, a reconstituted SLC5A8 complex comprises a reconstituted mixture of at least semi-purified proteins. By semi-purified, it is meant that the proteins utilized in the reconstituted mixture have been previously separated from

other cellular or viral proteins. For instance, in contrast to cell lysates, the proteins involved in SLC5A8 complex formation are present in the mixture to at least 50% purity relative to all other proteins in the mixture, and more preferably are present at 90-95% purity. In certain embodiments of the subject method, the reconstituted protein mixture is derived by mixing highly purified proteins such that the reconstituted mixture substantially lacks other proteins (such as of cellular or viral origin) which might interfere with or otherwise alter the ability to measure SLC5A8 complex assembly and/or disassembly.

Assaying SLC5A8 complexes, in the presence and absence of a candidate agent, can be accomplished in any vessel suitable for containing the reactants. Examples include microtitre plates, test tubes, and micro-centrifuge tubes. In a screening assay, the effect of a test agent may be assessed by, for example, assessing the effect of the test agent on kinetics, steady-state and/or endpoint of the reaction.

In one embodiment of the present invention, drug screening assays can be generated which detect inhibitory agents on the basis of their ability to interfere with assembly or stability of the SLC5A8 complex. In an exemplary binding assay, the compound of interest is contacted with a mixture comprising a SLC5A8 polypeptide and at least one interacting polypeptide. Detection and quantification of SLC5A8 complexes provides a means for determining the compound's efficacy at inhibiting (or potentiating) interaction between the two polypeptides. The efficacy of the compound can be assessed by generating dose response curves from data obtained using various concentrations of the test compound. Moreover, a control assay can also be performed to provide a baseline for comparison. In the control assay, the formation of complexes is quantitated in the absence of the test compound.

Complex formation between the SLC5A8 polypeptides and a substrate polypeptide may be detected by a variety of techniques. For instance, modulation in the formation of complexes can be quantitated using, for example, detectably labeled proteins (e.g., radiolabeled, fluorescently labeled, or enzymatically labeled), by immunoassay, or by chromatographic detection. Surface plasmon resonance systems, such as those available from Biacore International AB (Uppsala, Sweden), may also be used to detect protein-protein interaction.

Often, it will be desirable to immobilize one of the polypeptides to facilitate separation of complexes from uncomplexed forms of one of the proteins, as well as to accommodate automation of the assay. In an illustrative embodiment, a fusion protein can be provided which

adds a domain that permits the protein to be bound to an insoluble matrix. For example, GST-SLC5A8 fusion proteins can be adsorbed onto glutathione sepharose beads (Sigma Chemical, St. Louis, MO) or glutathione derivatized microtitre plates, which are then combined with a potential interacting protein, e.g., an ^{35}S-labeled polypeptide, and the test compound and incubated under conditions conducive to complex formation . Following incubation, the beads are washed to remove any unbound interacting protein, and the matrix bead-bound radiolabel determined directly (e.g., beads placed in scintillant), or in the supernatant after the complexes are dissociated, e.g., when microtitre plate is used. Alternatively, after washing away unbound protein, the complexes can be dissociated from the matrix, separated by SDS-PAGE gel, and the level of interacting polypeptide found in the matrix-bound fraction quantitated from the gel using standard electrophoretic techniques.

In a further embodiment, agents that bind to a SLC5A8 may be identified by using an immobilized SLC5A8. In an illustrative embodiment, a fusion protein can be provided which adds a domain that permits the protein to be bound to an insoluble matrix. For example, GST-SLC5A8 fusion proteins can be adsorbed onto glutathione sepharose beads (Sigma Chemical, St. Louis, MO) or glutathione derivatized microtitre plates, which are then combined with a potential labeled binding agent and incubated under conditions conducive to binding. Following incubation, the beads are washed to remove any unbound agent, and the matrix bead-bound label determined directly, or in the supernatant after the bound agent is dissociated.

In yet another embodiment, the SLC5A8 polypeptide and potential interacting polypeptide can be used to generate an interaction trap assay (see also, U.S. Patent No. 5,283,317; Zervos ct al. (1993) Cell 72:223-232; Madura et al. (1993) J Biol Chem 268:12046-12054; Bartel et al. (1993) Biotechniques 14:920-924; and Iwabuchi et al. (1993) Oncogene 8:1693-1696), for subsequently detecting agents which disrupt binding of the proteins to one and other.

One aspect of the present invention provides reconstituted protein preparations including a SLC5A8 polypeptide and one or more interacting polypeptides.

In still further embodiments of the present assay, the SLC5A8 complex is generated in whole cells, taking advantage of cell culture techniques to support the subject assay. For example, as described below, the SLC5A8 complex can be constituted in a eukaryotic cell culture system, including mammalian and yeast cells. Advantages to generating the subject

assay in an intact cell include the ability to detect inhibitors which are functional in an environment more closely approximating that which therapeutic use of the inhibitor would require, including the ability of the agent to gain entry into the cell. Furthermore, certain of the in vivo embodiments of the assay, such as examples given below, are amenable to high throughput analysis of candidate agents.

The components of the SLC5A8 complex can be endogenous to the cell selected to support the assay. Alternatively, some or all of the components can be derived from exogenous sources. For instance, fusion proteins can be introduced into the cell by recombinant techniques (such as through the use of an expression vector), as well as by microinjecting the fusion protein itself or mRNA encoding the fusion protein.

In many embodiments, a cell is manipulated after incubation with a candidate agent and assayed for a SLC5A8 activity. In certain embodiments a SLC5A8 activity is represented by sodium transporting activity or tumor suppressing activity. In certain embodiments, SLC5A8 activities may also include, without limitation, complex formation between SLC5A8 and its associated proteins. SLC5A8 complex formation may be assessed by immunoprecipitation and analysis of co-immunoprecipiated proteins or affinity purification and analysis of co-purified proteins. Fluorescence Resonance Energy Transfer (FRET)-based assays may also be used to determine complex formation. Fluorescent molecules having the proper emission and excitation spectra that are brought into close proximity with one another can exhibit FRET. The fluorescent molecules are chosen such that the emission spectrum of one of the molecules (the donor molecule) overlaps with the excitation spectrum of the other molecule (the acceptor molecule). The donor molecule is excited by light of appropriate intensity within the donor's excitation spectrum. The donor then emits the absorbed energy as fluorescent light. The fluorescent energy it produces is quenched by the acceptor molecule. FRET can be manifested as a reduction in the intensity of the fluorescent signal from the donor, reduction in the lifetime of its excited state, and/or re-emission of fluorescent light at the longer wavelengths (lower energies) characteristic of the acceptor. When the fluorescent proteins physically separate, FRET effects are diminished or eliminated. (U.S. Patent No. 5,981,200).

In general, where the screening assay is a binding assay (whether protein-protein binding, agent-protein binding, etc.), one or more of the molecules may be joined to a label, where the label can directly or indirectly provide a detectable signal. Various labels include radioisotopes, fluorescers, chemiluminescers, enzymes, specific binding molecules, particles,

e.g., magnetic particles, and the like. Specific binding molecules include pairs, such as biotin and streptavidin, digoxin and antidigoxin etc. For the specific binding members, the complementary member would normally be labeled with a molecule that provides for detection, in accordance with known procedures.

A variety of other reagents may be included in the screening assay. These include reagents like salts and neutral proteins (e.g., albumin, detergents, etc) that are used to facilitate optimal protein-protein binding and/or reduce nonspecific or background interactions. Reagents that improve the efficiency of the assay, such as protease inhibitors, nuclease inhibitors, anti-microbial agents, etc. may be used. The mixture of components are added in any order that provides for the requisite binding. Incubations are performed at any suitable temperature, typically between 4 °C and 40 °C. Incubation periods are selected for optimum activity, but may also be optimized to facilitate rapid high-throughput screening.

It is to be understood that the screening assays discussed above are applicable to identify therapeutic agents related to soluble SLC5A8 polypeptides and derivatives thereof. An exemplary derivative of soluble SLC5A8 polypeptides is a fusion protein containing soluble SLC5A8 polypeptide. Given the role of soluble SLC5A8 polypeptides in sodium transporting and/or tumor suppression, compositions that perturb the formation or stability of the protein-protein interactions between soluble SLC5A8 polypeptides and the proteins that they interact with, are candidate pharmaceuticals for the treatment of SLC5A8-associated diseases such as cancer.

IX. Predictive Medicine

The present invention also pertains to the field of predictive medicine in which diagnostic assays, prognostic assays, and monitoring clinical trials are used for prognostic (predictive) purposes to thereby treat an individual. Generally, the invention provides a method of determining if a subject is at risk for a disorder related to a lesion in or the misexpression of a gene which encodes SLC5A8, for example cancers (e.g., colon cancer, breast cancer, thyroid cancer, or stomach cancer).

The method includes one or more of the following: 1) detecting, in a tissue of the subject, the presence or absence of a mutation which affects the expression of the SLC5A8 gene, or detecting the presence or absence of a mutation in a region which controls the expression of the gene, e.g., a mutation in the 5' control region; 2) detecting, in a tissue of the subject, the

presence or absence of a mutation which alters the structure of the SLC5A8 gene; 3) detecting, in a tissue of the subject, the misexpression of the SLC5A8 gene, at the mRNA level, e.g., detecting a non-wild type level of a mRNA; 4) detecting, in a tissue of the subject, the misexpression of the gene, at the protein level, e.g., detecting a non-wild type level of a SLC5A8 polypeptide; and 5) detecting, in a tissue of the subject, methylation of the SLC5A8 gene in the 5' SLC5A8 genomic nucleotide sequences (see detailed descriptions in the following section).

In preferred embodiments, the method may also include ascertaining the existence of at least one of: 1) a deletion of one or more nucleotides from the SLC5A8 gene; 2) an insertion of one or more nucleotides into the gene; 3) a point mutation, e.g., a substitution of one or more nucleotides of the gene; and 4) a gross chromosomal rearrangement of the gene, e.g., a translocation, inversion, or deletion.

For example, detecting the genetic lesion can include: (i) providing a probe/primer including an oligonucleotide containing a region of nucleotide sequence which hybridizes to a sense or antisense sequence from SEQ ID NO: 3 or 4, or naturally occurring mutants thereof, or 5' or 3' flanking sequences naturally associated with the SLC5A8 gene; (ii) exposing the probe/primer to nucleic acid of the tissue; and detecting, by hybridization, e.g., in situ hybridization, of the probe/primer to the nucleic acid, the presence or absence of the genetic lesion.

In preferred embodiments, detecting the misexpression includes ascertaining the existence of at least one of: an alteration in the level of a messenger RNA transcript of the SLC5A8 gene; the presence of a non-wild type splicing pattern of a messenger RNA transcript of the gene; or a non-wild type level of SLC5A8.

Methods of the invention can be used prenatally or to determine if a subject's offspring will be at risk for a disorder. In preferred embodiments, the method includes determining the structure of a SLC5A8 gene, an abnormal structure being indicative of risk for the disorder.

In preferred embodiments, the method includes contacting a sample from the subject with an antibody to the SLC5A8 protein or a nucleic acid which hybridizes specifically with the gene. These and other embodiments are discussed below.

X. Diagnostic and Prognostic Assays

Diagnostic and prognostic assays of the invention include method for assessing the expression level of SLC5A8 molecules and for identifying variations and mutations in the sequence of SLC5A8 molecules. In certain embodiments, the invention provides methods by assaying the SLC5A8 expression level so as to determine whether a patient has or does not have a disease condition. Further, such a disease condition may be characterized by decreased expression of SLC5A8 nucleic acid or protein described herein. In certain embodiments, the invention provides methods for determining whether a patient is or is not likely to have a SLC5A8-associated disease by detecting the expression of the SLC5A8 nucleotide sequences. In further embodiments, the invention provides methods for determining whether the patient is having a relapse or determining whether a patient's cancer is responding to treatment.

The presence, level, or absence of SLC5A8 protein or nucleic acid in a biological sample can be evaluated by obtaining a biological sample from a test subject and contacting the biological sample with a compound or an agent capable of detecting SLC5A8 protein or nucleic acid (e.g., mRNA, genomic DNA) that encodes SLC5A8 protein such that the presence of SLC5A8 protein or nucleic acid is detected in the biological sample. The level of expression of the SLC5A8 gene can be measured in a number of ways, including, but not limited to: measuring the mRNA encoded by the SLC5A8 genes; measuring the amount of protein encoded by the SLC5A8 gene; or measuring the activity of the protein encoded by the SLC5A8 gene. The level of mRNA corresponding to the SLC5A8 gene in a cell can be determined both by in situ and by in vitro formats.

The isolated mRNA can be used in hybridization or amplification assays that include, but are not limited to, Southern or Northern analyses, polymerase chain reaction (PCR) analyses and probe arrays. One preferred diagnostic method for the detection of mRNA levels involves contacting the isolated mRNA with a nucleic acid molecule (probe) that can hybridize to the mRNA encoded by the SLC5A8 gene. The nucleic acid probe can be, for example, a full-length SLC5A8 nucleic acid, such as the nucleic acid of SEQ ID NO: 3 or 4, or a portion thereof, such as an oligonucleotide of at least 7, 15, 30, 50, 100, 250 or 500 nucleotides in length and sufficient to specifically hybridize under stringent conditions to SLC5A8 mRNA or genomic DNA. The probe can be disposed on an address of an array, e.g., an array described below. Other suitable probes for use in the diagnostic assays are described herein.

In one format, mRNA (or cDNA) is immobilized on a surface and contacted with the probes, for example, by running the isolated mRNA on an agarose gel and transferring the

mRNA from the gel to a membrane, such as nitrocellulose. In an alternative format, the probes are immobilized on a surface and the mRNA (or cDNA) is contacted with the probes, for example, in a two-dimensional gene chip array described below. A skilled artisan can adapt known mRNA detection methods for use in detecting the level of mRNA encoded by the SLC5A8 gene.

The level of SLC5A8 mRNA in a sample can be evaluated with nucleic acid amplification, e.g., by RT-PCR (Mullis (1987) U.S. Pat. No. 4,683,202), ligase chain reaction (Barany (1991) Proc. Natl. Acad. Sci. USA 88:189-193), self sustained sequence replication (Guatelli et al., (1990) Proc. Natl. Acad. Sci. USA 87:1874-1878), transcriptional amplification system (Kwoh et al., (1989), Proc. Natl. Acad. Sci. USA 86:1173-1177), Q-Beta Replicase (Lizardi et al., (1988) Bio/Technology 6:1197), rolling circle replication (Lizardi et al., U.S. Patent. No. 5,854,033) or any other nucleic acid amplification method, followed by the detection of the amplified molecules using techniques known in the art. As used herein, amplification primers are defined as being a pair of nucleic acid molecules that can anneal to 5' or 3' regions of a gene (plus and minus strands, respectively, or vice-versa) and contain a short region in between. In general, amplification primers are from about 10 to 30 nucleotides in length and flank a region from about 50 to 200 nucleotides in length. Under appropriate conditions and with appropriate reagents, such primers permit the amplification of a nucleic acid molecule comprising the nucleotide sequence flanked by the primers.

For in situ methods, a cell or tissue sample can be prepared/processed and immobilized on a support, typically a glass slide, and then contacted with a probe that can hybridize to mRNA that encodes the SLC5A8 gene being analyzed.

In another embodiment, the methods further contacting a control sample with a compound or agent capable of detecting SLC5A8 mRNA, or genomic DNA, and comparing the presence of SLC5A8 mRNA or genomic DNA in the control sample with the presence of SLC5A8 mRNA or genomic DNA in the test sample.

A variety of methods can be used to determine the level of protein encoded by SLC5A8. In general, these methods include contacting an agent that selectively binds to the protein, such as an antibody with a sample, to evaluate the level of protein in the sample. In a preferred embodiment, the antibody bears a detectable label. Antibodies can be polyclonal, or more preferably, monoclonal. An intact antibody, or a fragment thereof (e.g., Fab or $F(ab')_2$) can be

used. The term "labeled," with regard to the probe or antibody, is intended to encompass direct labeling of the probe or antibody by coupling (i.e., physically linking) a detectable substance to the probe or antibody, as well as indirect labeling of the probe or antibody by reactivity with a detectable substance. Examples of detectable substances are provided herein.

The detection methods can be used to detect SLC5A8 protein in a biological sample in vitro as well as in vivo. In vitro techniques for detection of SLC5A8 protein include enzyme linked immunosorbent assays (ELISAs), immunoprecipitations, immunofluorescence, enzyme immunoassay (EIA), radioimmunoassay (RIA), and Western blot analysis. In vivo techniques for detection of SLC5A8 protein include introducing into a subject a labeled anti-SLC5A8 antibody. For example, the antibody can be labeled with a radioactive marker whose presence and location in a subject can be detected by standard imaging techniques. In another embodiment, the sample is labeled, e.g., biotinylated and then contacted to the antibody, e.g., an anti-SLC5A8 antibody positioned on an antibody array (as described below). The sample can be detected, e.g., with avidin coupled to a fluorescent label.

In another embodiment, the methods further include contacting the control sample with a compound or agent capable of detecting SLC5A8 protein, and comparing the presence of SLC5A8 protein in the control sample with the presence of SLC5A8 protein in the test sample.

The invention also includes kits for detecting the presence of SLC5A8 in a biological sample. For example, the kit can include a compound or agent capable of detecting SLC5A8 protein or mRNA in a biological sample; and a standard. The compound or agent can be packaged in a suitable container. The kit can further comprise instructions for using the kit to detect SLC5A8 protein or nucleic acid.

For antibody-based kits, the kit can include: (1) a first antibody (e.g., attached to a solid support) which binds to a polypeptide corresponding to a marker of the invention; and, optionally, (2) a second, different antibody which binds to either the polypeptide or the first antibody and is conjugated to a detectable agent.

For oligonucleotide-based kits, the kit can include: (1) an oligonucleotide, e.g., a detectably labeled oligonucleotide, which hybridizes to a nucleic acid sequence encoding a polypeptide corresponding to a marker of the invention or (2) a pair of primers useful for amplifying a nucleic acid molecule corresponding to a marker of the invention. The kit can also

buffering agent, a preservative, or a protein stabilizing agent. The kit can also

includes components necessary for detecting the detectable agent (e.g., an enzyme or a substrate). The kit can also contain a control sample or a series of control samples which can be assayed and compared to the test sample contained. Each component of the kit can be enclosed within an individual container and all of the various containers can be within a single package, along with instructions for interpreting the results of the assays performed using the kit.

The diagnostic methods described herein can identify subjects having, or at risk of developing, a disease or disorder associated with misexpressed or aberrant or unwanted SLC5A8 expression or activity. As used herein, the term "unwanted" includes an unwanted phenomenon involved in a biological response such as pain or deregulated cell proliferation.

In one embodiment, a disease or disorder associated with aberrant or unwanted SLC5A8 expression or activity is identified. A test sample is obtained from a subject and SLC5A8 protein or nucleic acid (e.g., mRNA or genomic DNA) is evaluated, wherein the level, e.g., the presence or absence, of SLC5A8 protein or nucleic acid is diagnostic for a subject having or at risk of developing a disease or disorder associated with aberrant or unwanted SLC5A8 expression or activity.

The prognostic assays described herein can be used to determine whether a subject can be administered an agent (e.g., an agonist, antagonist, peptidomimetic, protein, peptide, nucleic acid, small molecule, or other drug candidate) to treat a disease or disorder associated with aberrant or unwanted SLC5A8 expression or activity. For example, such methods can be used to determine whether a subject can be effectively treated with an agent for a pain or solute transport disorder.

In yet another aspect, the invention features a method of evaluating a test compound (see also, "Screening Assays", above). The method includes providing a cell and a test compound; contacting the test compound to the cell; obtaining a subject expression profile for the contacted cell; and comparing the subject expression profile to one or more reference profiles. The profiles include a value representing the level of SLC5A8 expression. In a preferred embodiment, the subject expression profile is compared to a target profile, e.g., a profile for a normal cell or for desired condition of a cell. The test compound is evaluated favorably if the subject expression profile is more similar to the target profile than an expression profile obtained from an uncontacted cell.

-- - - thods of Assaying Methylation of SLC5A8 Nucleotides

In certain aspects, the invention provides assays and methods using the SLC5A8 nucleotide sequences as molecular markers that distinguish between healthy cells and SLC5A8-associated diseased cells (cells of colon cancer, breast cancer, thyroid cancer or stomach cancer). In one aspect, a molecular marker of the invention is a differentially methylated SLC5A8 nucleotide sequence.

Accordingly, in certain embodiments, the invention provides assays for detecting differentially methylated SLC5A8 nucleotide sequences, such as the differential methylation patterns in nucleic acid sequence of SEQ ID NO: 12, 13 or 14. Thus, a differentially methylated SLC5A8 nucleotide sequence, in its methylated state, can be a SLC5A8-associated cancer-specific modification that serves as a target for detection using various methods described herein and the methods that are well within the purview of the skilled artisan in view of the teachings of this application.

In certain aspects, such methods for detecting methylated SLC5A8 nucleotide sequences are based on treatment of SLC5A8 genomic DNA with a chemical compound which converts non-methylated C, but not methylated C (i.e., 5mC), to a different nucleotide base. One such compound is sodium bisulfite, which converts C, but not 5mC, to U. Methods for bisulfite treatment of DNA are known in the art (Herman, et al., 1996, Proc Natl Acad Sci USA, 93:9821-6; Herman and Baylin, 1998, Current Protocols in Human Genetics, N. E. A. Dracopoli, ed., John Wiley & Sons, 2:10.6.1-10.6.10; U.S. Patent No. 5,786,146). To illustrate, when an DNA molecule that contains unmethylated C nucleotides is treated with sodium bisulfite to become a compound-converted DNA, the sequence of that DNA is changed (C→U). Detection of the U in the converted nucleotide sequence is indicative of an unmethylated C.

The different nucleotide base (e.g., U) present in compound-converted nucleotide sequences can subsequently be detected in a variety of ways. In a preferred embodiment, the present invention provides a method of detecting U in compound-converted SLC5A8 DNA sequences by using "methylation sensitive PCR" (MSP) (see, e.g., Herman, et al., 1996, *Proc. Natl. Acad. Sci. USA*, 93:9821-9826; U.S. Patent Nos. 6,265,171; 6,017,704; and 6,200,756). In MSP, one set of primers (i.e., comprising a forward and a reverse primer) amplifies the compound-converted template sequence if C bases in CpG dinucleotides within the SLC5A8 DNA are methylated. This set of primers is called "methylation-specific primers." Another set of primers amplifies the compound-converted template sequence if C bases in CpG

dinucleotides within the SLC5A8 5' flanking sequence are not methylated. This set of primers is called "unmethylation-specific primers."

In MS-PCR, the reactions use the compound-converted DNA from a sample in a subject. In assays for SLC5A8 methylated DNA, methylation-specific primers are used. In the case where C within CpG dinucleotides of the target sequence of the DNA are methylated, the methylation-specific primers will amplify the compound-converted template sequence in the presence of a polymerase and an MSP product will be produced. If C within CpG dinucleotides of the target sequence of the DNA are not methylated, the methylation-specific primers will not amplify the compound-converted template sequence in the presence of a polymerase and an MSP product will not be produced

It is often also useful to run a control reaction for the detection of unmethylated SLC5A8 DNA. The reactions uses the compound-converted DNA from a sample in a subject and unmethylation-specific primers are used. In the case where C within CpG dinucleotides of the target sequence of the DNA are unmethylated, the unmethylation specific primers will amplify the compound-converted template sequence in the presence of a polymerase and an MSP product will be produced. If C within CpG dinucleotides of the target sequence of the DNA are methylated, the unmethylation-specific primers will not amplify the compound-converted template sequence in the presence of a polymerase and an MSP product will not be produced. Note that a biologic sample will often contain a mixture of both neoplastic cells that give rise to a signal with methylation specific primers, and normal cellular elements that give rise to a signal with unmethylation-specific primers. The unmethyl specific signal is often of use as a control reaction, but does not in this instance imply the absence of cancer (e.g., colon cancer, breast cancer, thyroid cancer, or stomach cancer) as indicated by the positive signal derived from reactions using the methylation specific primers.

Primers for an MSP reaction are derived from the compound-converted SLC5A8 template sequence. Herein, "derived from" means that the sequences of the primers are chosen such that the primers amplify the compound-converted template sequence in an MSP reaction. Each primer comprises a single-stranded DNA fragment which is at least 8 nucleotides in length. Preferably, the primers are less than 50 nucleotides in length, more preferably from 15 to 35 nucleotides in length. Because the compound-converted SLC5A8 template sequence can be either the Watson strand or the Crick strand of the double-stranded DNA that is treated with [illegible] bisulfite, the sequences of the primers is dependent upon whether the Watson or Crick

compound-converted template sequence is chosen to be amplified in the MSP. Either the Watson or Crick strand can be chosen to be amplified.

The compound-converted SLC5A8 template sequence, and therefore the product of the MSP reaction, can be between 20 to 3000 nucleotides in length, preferably between 50 to 500 nucleotides in length, more preferably between 80 to 150 nucleotides in length. Preferably, the methylation-specific primers result in an MSP product of a different length than the MSP product produced by the unmethylation-specific primers.

A variety of methods can be used to determine if an MSP product has been produced in a reaction assay. One way to determine if an MSP product has been produced in the reaction is to analyze a portion of the reaction by agarose gel electrophoresis. For example, a horizontal agarose gel of from 0.6 to 2.0% agarose is made and a portion of the MSP reaction mixture is electrophoresed through the agarose gel. After electrophoresis, the agarose gel is stained with ethidium bromide. MSP products are visible when the gel is viewed during illumination with ultraviolet light. By comparison to standardized size markers, it is determined if the MSP product is of the correct expected size.

Other methods can be used to determine whether a product is made in an MSP reaction. One such method is called "real-time PCR." Real-time PCR utilizes a thermal cycler (i.e., an instrument that provides the temperature changes necessary for the PCR reaction to occur) that incorporates a fluorimeter (i.e. an instrument that measures fluorescence). The real-time PCR reaction mixture also contains a reagent whose incorporation into a product can be quantified and whose quantification is indicative of copy number of that sequence in the template. One such reagent is a fluorescent dye, called SYBR Green I (Molecular Probes, Inc.; Eugene, Oregon) that preferentially binds double-stranded DNA and whose fluorescence is greatly enhanced by binding of double-stranded DNA. When a PCR reaction is performed in the presence of SYBR Green I, resulting DNA products bind SYBR Green I and fluorescence. The fluorescence is detected and quantified by the fluorimeter. Such technique is particularly useful for quantification of the amount of the product in the PCR reaction. Additionally, the product from the PCR reaction may be quantitated in "real-time PCR" by the use of a variety of probes that hybridize to the product including TaqMan probes and molecular beacons. Quantitation may be on an absolute basis, or may be relative to a constitutively methylated DNA standard, or may be relative to an unmethylated DNA standard. In one instance the ratio of methylated SLC5A8 derived product to unmethylated derived SLC5A8 product may be constructed.

Methods for detecting methylation of the SLC5A8 DNA in this invention are not limited to MSP, and may cover any assay for detecting DNA methylation. Another example method for detecting methylation of the SLC5A8 DNA is by using "methylation-sensitive" restriction endonucleases. Such methods comprise treating the genomic DNA isolated from a subject with an methylation-sensitive restriction endonuclease and then using the restriction endonuclease-treated DNA as a template in a PCR reaction. Herein, methylation-sensitive restriction endonucleases recognize and cleave a specific sequence within the DNA if C bases within the recognition sequence are not methylated. If C bases within the recognition sequence of the restriction endonuclease are methylated, the DNA will not be cleaved. Examples of such methylation-sensitive restriction endonucleases include, but are not limited to HpaII, SmaI, SacII, EagI, MspI, BstUI, and BssHII. In this technique, a recognition sequence for a methylation-sensitive restriction endonuclease is located within the template DNA, at a position between the forward and reverse primers used for the PCR reaction. In the case that a C base within the methylation-sensitive restriction endonuclease recognition sequence is not methylated, the endonuclease will cleave the DNA template and a PCR product will not be formed when the DNA is used as a template in the PCR reaction. In the case that a C base within the methylation-sensitive restriction endonuclease recognition sequence is methylated, the endonuclease will not cleave the DNA template and a PCR product will be formed when the DNA is used as a template in the PCR reaction. Therefore, methylation of C bases can be determined by the absence or presence of a PCR product (Kane, et al., 1997, Cancer Res, 57:808-11). No sodium bisulfite is used in this technique.

Yet another exemplary method for detecting methylation of the SLC5A8 DNA is called the modified MSP, which method utilizes primers that are designed and chosen such that products of the MSP reaction are susceptible to digestion by restriction endonucleases, depending upon whether the compound-converted template sequence contains CpG dinucleotides or UpG dinucleotides.

Yet other methods for detecting methylation of the SLC5A8 DNA include the MS-SnuPE methods. This method uses compound-converted SLC5A8 DNA as a template in a primer extension reaction wherein the primers used produce a product, dependent upon whether the compound-converted template contains CpG dinucleotides or UpG dinucleotides (see e.g., Gonzalgo, et al., 1997, *Nucleic Acids Res.*, 25:2529-31).

Another exemplary method for detecting methylation of the SLC5A8 DNA is called COBRA (i.e., combined bisulfite restriction analysis). This method has been routinely used for DNA methylation detection and is well known in the art (see, e.g., Xiong, et al., 1997, *Nucleic Acids Res*, 25:2532-4).

In certain embodiments, the invention provides methods that involve directly sequencing the product resulting from an MSP reaction to determine if the compound-converted SLC5A8 template sequence contains CpG dinucleotides or UpG dinucleotides. Molecular biology techniques such as directly sequencing a PCR product are well known in the art.

XII .SLC5A8 Oligonucleotides for Methylation Detection

In yet other aspects, the application provides oligonucleotide primers for amplifying a region within the SLC5A8 nucleic acid sequence of any one of SEQ ID NOs: 5-11. In certain aspects, a pair of the oligonucleotide primers (for example, SEQ ID NOs: 5-7) can be used in a detection assay, such as the HpaII assay. In certain aspects, primers used in an MSP reaction can specifically distinguish between methylated and non-methylated SLC5A8 DNA, for example, SEQ ID NOs: 8-11.

The primers of the invention have sufficient length and appropriate sequence so as to provide specific initiation of amplification of SLC5A8 nucleic acids. Primers of the invention are designed to be "substantially" complementary to each strand of the SLC5A8 nucleic acid sequence to be amplified. While exemplary primers are provided in SEQ ID NOs: 5-11, it is understood that any primers that hybridizes with the bisulfite-converted SLC5A8 sequence of SEQ ID NOs: 12-14 are included within the scope of this invention and is useful in the method of the invention for detecting methylated nucleic acid, as described. Similarly, it is understood that any primers that would serve to amplify a methylation sensitive restriction site or sites within the differentially methylated region of SEQ ID NOs: 12-14 are included within the scope of this invention and is useful in the method of the invention for detecting nucleic methylated nucleic acid, as described.

The oligonucleotide primers of the invention may be prepared by using any suitable method, such as conventional phosphotriester and phosphodiester methods or automated embodiments thereof. In one such automated embodiment, diethylphosphoramidites are used as starting materials and may be synthesized as described by Beaucage, et al. (*Tetrahedron Letters*,

22:1859-1862, 1981). One method for synthesizing oligonucleotides on a modified solid support is described in U.S. Patent No. 4,458,066.

In particular, a pair of primers are selected to amplify the SLC5A8 methylation target region or a DNA segment thereof. The targeted DNA segment that is amplified by the primers contains a plurality of sites that are recognized by the methylation sensitive restriction enzyme and is located between base pairs 82200 to 83267 of GenBank entry AC063951. In one preferred embodiment, the targeted DNA segment comprises at least four HpaII sites and the primers amplify a region including base pair 82638 through base pair 83080 of GenBank entry AC063951. In another highly preferred embodiment, the targeted DNA segment comprises at least six HpaII sites and the primers amplify a region including base pair 82430 through base pair 83080 of GenBank entry AC063951.

For example, each primer comprises a single-stranded DNA fragment which is at least 8 nucleotides in length. Preferably, the primers are less than 50 nucleotides in length, more preferably from 15 to 35 nucleotides in length. The sequences of the primers are derived from the sequence of the targeted DNA segment, i.e., the segment that is to be amplified. The sequence of the forward primer is identical to a sequence at the 5' end of the targeted DNA segment. The sequence of the reverse primer is the reverse complement of a sequence at the 3' end of targeted DNA segment.

XIII. Subjects and Samples

In certain aspects, the invention relates to a subject suspected of having or has a SLC5A8-associated disease, such as colon cancer, breast cancer, thyroid cancer, or stomach cancer. Alternatively, a subject may be undergoing routine screening and may not necessarily be suspected of having such a SLC5A8-associated disease or condition. In a preferred embodiment, the subject is a human subject, and the SLC5A8-associated disease is colon neoplasia.

Assaying for SLC5A8 markers discussed above in a sample from subjects not known to have a cancer (e.g., colon cancer, breast cancer, thyroid cancer, or stomach cancer) can aid in diagnosis of such a cancer in the subject. To illustrate, detecting the methylation status of the SLC5A8 nucleotide sequence by MSP can be used by itself, or in combination with other various assays, to improve the sensitivity and/or specificity for detecting a cancer. Preferably,

such a detection is made at an early stage in the development of cancer, so that treatment is more likely to be effective.

In addition to diagnosis, assaying of a SLC5A8 marker in a sample from a subject not known to have a cancer (e.g., colon cancer, breast cancer, thyroid cancer, or stomach cancer) can be prognostic for the subject (e.g., indicating the probable course of the disease). To illustrate, subjects having a predisposition to develop colon neoplasia may possess methylated SLC5A8 nucleotide sequences. Assaying of SLC5A8 markers in a samples from subjects can also be used to select a particular therapy or therapies which are particularly effective against the colon neoplasia in the subject, or to exclude therapies that are not likely to be effective.

Assaying of SLC5A8 markers in samples from subjects that are known to have, or to have had, a cancer associated with silencing of the SLC5A8 gene is also useful. For example, the present methods can be used to identify whether therapy is effective or not for certain subjects. One or more samples are taken from the same subject prior to and following therapy, and assayed for the SLC5A8 markers. A finding that the SLC5A8 marker is present in the sample taken prior to therapy and absent (or at a lower level) after therapy would indicate that the therapy is effective and need not be altered. In those cases where the SLC5A8 marker is present in the sample taken before therapy and in the sample taken after therapy, it may be desirable to alter the therapy to increase the likelihood that the cancer will be eradicated in the subject. Thus, the present method may obviate the need to perform more invasive procedures which are used to determine a patient's response to therapy.

Cancers frequently recur following therapy in patients with advanced cancers. In this and other instances, the assays of the invention are useful for monitoring over time the status of an cancer associated with silencing of the SLC5A8 gene. For subjects in which a cancer is progressing, a SLC5A8 marker may be absent from some or all samples when the first sample is taken and then appear in one or more samples when the second sample is taken. For subjects in which cancer is regressing, a SLC5A8 marker may be present in one or a number of samples when the first sample is taken and then be absent in some or all of these samples when the second sample is taken.

Samples for use with the methods described herein may be essentially any biological material of interest. For example, a sample may be a bodily fluid sample from a subject, a tissue sample from a subject, a solid or semi-solid sample from a subject, a primary cell culture or

tissue culture of materials derived from a subject, cells from a cell line, or medium or other extracellular material from a cell or tissue culture, or a xenograft (meaning a sample of a cancer from a first subject, e.g., a human, that has been cultured in a second subject, e.g., an immuno-compromised mouse). The term "sample" as used herein is intended to encompass both a biological material obtained directly from a subject (which may be described as the primary sample) as well as any manipulated forms or portions of a primary sample. A sample may also be obtained by contacting a biological material with an exogenous liquid, resulting in the production of a lavage liquid containing some portion of the contacted biological material. Furthermore, the term "sample" is intended to encompass the primary sample after it has been mixed with one or more additive, such as preservatives, chelators, anti-clotting factors, etc.

In certain embodiments, a bodily fluid sample is a blood sample. In this case, the term "sample" is intended to encompass not only the blood as obtained directly from the patient but also fractions of the blood, such as plasma, serum, cell fractions (e.g., platelets, erythrocytes, and lymphocytes), protein preparations, nucleic acid preparations, etc. In certain embodiments, a bodily fluid sample is a urine sample or a colonic effluent sample. In certain embodiments, a bodily fluid sample is a stool sample.

A subject is preferably a human subject, but it is expected that the molecular markers disclosed herein, and particularly their homologs from other animals, are of similar utility in other animals. In certain embodiments, it may be possible to detect a SLC5A8 marker directly in an organism without obtaining a separate portion of biological material. In such instances, the term "sample" is intended to encompass that portion of biological material that is contacted with a reagent or device involved in the detection process.

In certain embodiments, DNA which is used as the template in an MSP reaction is obtained from a bodily fluid sample. Examples of preferred bodily fluids are blood, serum, plasma, a blood-derived fraction, stool, colonic effluent or urine. Other body fluids can also be used. Because they can be easily obtained from a subject and can be used to screen for multiple diseases, blood or blood-derived fractions are especially useful. For example, it has been shown that DNA alterations in colorectal cancer patients can be detected in the blood of subjects (Hibi, et al., 1998, Cancer Res, 58:1405-7). Blood-derived fractions can comprise blood, serum, plasma, or other fractions. For example, a cellular fraction can be prepared as a "buffy coat" (i.e., leukocyte-enriched blood portion) by centrifuging 5 ml of whole blood for 10 min at 800 times gravity at room temperature. Red blood cells sediment most rapidly and are present as the

bottom-most fraction in the centrifuge tube. The buffy coat is present as a thin creamy white colored layer on top of the red blood cells. The plasma portion of the blood forms a layer above the buffy coat. Fractions from blood can also be isolated in a variety of other ways. One method is by taking a fraction or fractions from a gradient used in centrifugation to enrich for a specific size or density of cells.

DNA is then isolated from samples from the bodily fluids. Procedures for isolation of DNA from such samples are well known to those skilled in the art. Commonly, such DNA isolation procedures comprise lysis of any cells present in the samples using detergents, for example. After cell lysis, proteins are commonly removed from the DNA using various proteases. RNA is removed using RNase. The DNA is then commonly extracted with phenol, precipitated in alcohol and dissolved in an aqueous solution.

XIV. Therapeutic methods for SLC5A8-associated diseases.

Yet another aspect of this application pertains to methods of treating a SLC5A8-associated disease (e.g., a proliferative disease such as cancer) which arises from reduced expression or over-expression of the SLC5A8 gene in cells. In certain cases, such SLC5A8-associated diseases (for example, colon cancer, breast cancer, thyroid cancer, or stomach cancer) can result from a wide variety of pathological cell proliferative conditions. In certain embodiments, treatment of a SLC5A8-associated disorder includes modulation of the SLC5A8 gene expression or SLC5A8 activity. The term "modulate" envisions the suppression of expression of SLC5A8 when it is over-expressed, or augmentation of SLC5A8 expression when it is under-expressed.

In an embodiment, the present invention provides a therapeutic method by using a SLC5A8 gene construct as a part of a gene therapy protocol, such as to reconstitute the function of a SLC5A8 protein (e.g., SEQ ID NO: 1) in a cell in which the SLC5A8 protein is mis-expressed or non-expressed. To illustrate, cell types which exhibit pathological or abnormal growth presumably depend at least in part on a function of a SLC5A8 protein. For example, gene therapy constructs encoding the SLC5A8 protein can be utilized in a cancer that is associated with silencing of the SLC5A8 gene, such as colon cancer, breast cancer, thyroid cancer, or stomach cancer.

In certain embodiments, the invention provides therapeutic methods using agents which xpression of SLC5A8. Loss of SLC5A8 gene expression in a SLC5A8-associated

diseased cells may be due at least in part to methylation of the SLC5A8 nucleotide sequence, methylation suppressive agents such as 5-deoxyazacytidine or 5-azacytidine can be introduced into the diseased cells. Other similar agents will be known to those of skill in the art. In a preferred embodiment, the SLC5A8-associated disease is colon neoplasia associated with increased methylation of SLC5A8 nucleotide sequences.

The present invention also provides gene therapy for the treatment of proliferative or immunologic disorders which are associated with SLC5A8. Such therapy would achieve its therapeutic effect by introduction of the SLC5A8 polynucleotide encoding full-length SLC5A8 into diseased cells.

Delivery of the SLC5A8 polynucleotide or the SLC5A8 gene can be achieved using a recombinant expression vector such as a chimeric virus or a colloidal dispersion system. Especially preferred for therapeutic delivery of antisense sequences is the use of targeted liposomes. Various viral vectors which can be utilized for gene therapy as taught herein include adenovirus, herpes virus, vaccinia, or, preferably, an RNA virus such as a retrovirus. Preferably, the retroviral vector is a derivative of a murine or avian retrovirus. Examples of retroviral vectors in which a single foreign gene can be inserted include, but are not limited to: Moloney murine leukemia virus (MoMuLV), Harvey murine sarcoma virus (HaMuSV), murine mammary tumor virus (MuMTV), and Rous Sarcoma Virus (RSV). Preferably, when the subject is a human, a vector such as the gibbon ape leukemia virus (GaLV) is utilized. A number of additional retroviral vectors can incorporate multiple genes. All of these vectors can transfer or incorporate a gene for a selectable marker so that transduced cells can be identified and generated. By inserting a SLC5A8 sequence of interest into the viral vector, along with another gene which encodes the ligand for a receptor on a specific target cell, for example, the vector is target-specific. Retroviral vectors can be made target-specific by attaching, for example, a sugar, a glycolipid or a protein. Preferred targeting is accomplished by using an antibody to target the retroviral vector. Those skilled in the art will know of, or can readily ascertain without undue experimentation, specific polynucleotide sequences which can be inserted into the retroviral genome or attached to a viral envelope to allow target-specific delivery of the retroviral vector containing the SLC5A8 gene.

The invention also relates to a medicament or pharmaceutical composition comprising a SLC5A8 5' flanking polynucleotide or a SLC5A8 5' flanking polynucleotide operably linked to the SLC5A8 structural gene, respectively, in a pharmaceutically acceptable excipient or medium

wherein the medicament is used for therapy of SLC5A8-associated diseases, such as colon cancer, breast cancer, thyroid cancer, or stomach cancer.

Exemplification

The invention now being generally described, it will be more readily understood by reference to the following examples, which are included merely for purposes of illustration of certain aspects and embodiments of the present invention, and are not intended to limit the invention.

Abstract:

We identify a new gene, *SLC5A8*, and show it is a candidate tumor suppressor gene whose silencing by aberrant methylation is a common and early event in human colon neoplasia. Aberrant DNA methylation has been implicated as a component of an epigenetic mechanism that silences genes in human cancers. Using restriction landmark genome scanning, we performed a global search to identify new genes that would be aberrantly methylated at high frequency in human colon cancer. From among 1,231 genomic NotI sites assayed, site 3D41 was identified as methylated in 11 of 12 colon cancers profiled. Site 3D41 mapped to exon 1 of *SLC5A8*, a novel transcript that we assembled. In normal colon mucosa we found *SLC5A8* exon 1 is unmethylated, and *SLC5A8* transcript is expressed. In contrast, *SLC5A8* exon 1 proved aberrantly methylated in 59% of primary colon cancers and 52% of colon cancer cell lines. *SLC5A8* exon 1 methylated cells were uniformly silenced for SLC5A8 expression, but reactivated expression upon treatment with a demethylating drug, 5-azacytidine. Transfection of *SLC5A8* suppressed colony growth in each of three SLC5A8 deficient cell lines, but showed no suppressive effect in any of three SLC5A8 proficient cell lines. *SLC5A8* exon 1 methylation is an early event, detectable in colon adenomas, and in even earlier microscopic colonic aberrant crypt foci. Structural homology and functional testing demonstrated SLC5A8 is a novel member of the family of sodium solute symporters, which are now added as a new class of candidate colon cancer suppressor genes.

Introduction:

Cytosine methylation within CpG dinucleotides is a recognized epigenetic DNA modification, which in normal human tissues is excluded from CpG rich "islands" that mark the promoters of certain genes (Baylin, et al., 1998, Adv Cancer Res 72:141-96; Jones, et al., 1999,

Trends Genet 15: 34-7; Baylin, et al., 2002, Cancer Cell 1: 299-305). Global hypomethylation accompanied by aberrant focal CpG island hypermethylation has emerged as one of the signature alterations evidenced by the cancer genome (Baylin, et al., 1998, Adv Cancer Res 72:141-96; Jones, et al., 1999, Trends Genet 15:34-7; Baylin, et al., 2002, Cancer Cell 1:299-305; Feinberg, et al., 1983, Nature 301:89-92). Moreover, silencing of gene expression as marked by aberrant methylation of CpG island promoter regions has emerged as a novel mechanism for the inactivation of tumor suppressor genes that provides an alternative to either mutation or to allelic loss (Baylin, et al., 1998, Adv Cancer Res 72:141-96; Jones, et al., 1999, Trends Genet 15:34-7; Kane, et al., 1997, Cancer Res 57:808-11; Veigl, et al., 1998, Proc Natl Acad Sci U S A 95:8698-702). Additionally, aberrant methylation of defined genomic sequences can serve as a potentially useful diagnostic marker for detection of human cancers (Grady, et al., 2001, Cancer Res 61:900-2; Usadel, et al., 2002, Cancer Res 62S:371-5).

Restriction landmark genome scanning (RLGS) provides a global analysis of methylation events in a cancer cell by providing a two dimensional display of the methylation status of genomic Not1 sites (Costello, et al., 2000, Nat Genet 24:132-8). To identify new tumor suppressor genes and /or identify new genes targeted for methylation in human colon cancer, we carried out RLGS analysis of 12 colon cancer cell lines. This analysis lead to the identification of a novel transcript *SLC5A8*, whose aberrant methylation and transcriptional silencing was found to be a common and early event in human colon cancers, and that was found to encode a novel sodium symporter whose restoration can markedly suppress colony forming ability of colon cells in which endogenous *SLC5A8* has been inactivated.

Significance:

This study demonstrates the application of restriction landmark genome scanning to identify a novel high frequency aberrant methylation event in human colon cancer. We extend that observation to identify a novel sodium transporter, SLC5A8, silenced by the methylation event. SLC5A8 methylation is among the most frequent molecular alterations in colon cancer, and finding SLC5A8 is a growth suppressor adds sodium transporters as a new functional class that can act as tumor suppressors. Moreover, detecting SLC5A8 methylation in aberrant crypt foci demonstrates this event as one of the earliest molecular changes in colon neoplasia, and adds further molecular support to the model in which at least some aberrant crypt foci are able to progress to more advanced colon adenomas and cancers.

Example 1:

Figure 3 depicts certain aspects of the present invention. The numerical coordinates are those of genomic clone AC063951. Lollipops designate CpG sites that are potential acceptors of aberrant methylation. Asterisks designate sites recognized by the HpaII restriction enzyme that cut these sites if unmethylated, but not if methylated. Shown are the positions of PCR primers that amplify regions crossing 6 HpaII sites, or regions crossing 4 HpaII sites. Also shown is the position of PCR primers designed for a methyl-specific PCR (MS-PCR) assays that amplify sodium bisulfite converted DNA specifically derived from templates that are either methylated or unmethylated at CpG dinucleotides interrogated by the PCR primers. Also shown in the gray bar is the 5' end of exon 1 of the SLC5A8 transcript which overlaps with the methylation sites detected in both MS-PCR and HpaII based assays. Lastly indicated is a site corresponding to methylation site 2D41 detected in Restriction Landmark Genome Scanning assay as methylated in colon cancer cell lines, though not in primary tumors.

Colon cancers that are aberrantly methylated can be detected as they are resistant to cutting by the HpaII enzyme. That is methylation in a colon cancer can be assayed by showing PCR amplification of a DNA product using the primers and conditions shown from DNA that has first been digested with the HpaII restriction enzyme. The assay is diagrammed in Figure 4 that provides the sequence of AC063951 between base pairs 82200-83267, and designates every CpG site with a gray lollipop, and shows the HpaII sites in the assay as black lollipops, and also shows the location of the PCR primers used in this assay. In this figure, the base pairs have been renumbered sequentially from 1-1068, with basepair 82200 being renumbered as basepair 1.

Figure 5 tabulates the correspondence of assay for methylation over 4 and 6 HpaII sites with silencing of expression of the SLC5A8 transcript. As noted, assay of methylation over 4 HpaII sites detects 100% of colon cancer cell lines that silence the SLC5A8 transcript, but also detects some colon cancer cell lines that express SLC5A8. Assay of methylation over 6 HpaII sites has 100% specificity and detects only cell lines that have silenced SLC5A8, with a sensitivity of 68%.

Figure 6 tabulates the results of this assay in actual colon cancer tumors. In a group of 34 human colon cancers 76% are detected by resistance to cutting at 4 HpaII sites whereas 50% are detected by resistance to cutting at 6 HpaII sites. Both assays detect methylation in some normal tissues accompanying methylated cancers, suggesting the detection of microscopic colon

cancer cells. No methylation is detected in any normal tissue in which the accompanying tumor is unmethylated. Because of its high specificity, the assay which employs methylation over 6 HpaII sites is preferred.

Figure 7 shows the results of assay for methylation at 61 CpG sites enumerated in Figure 4 with site 1 corresponding to basepair 466 in Figure 4 and site 61 corresponding to basepair 1010. The bold arrows correspond to 4 of the HpaII sites at respectively basepairs 466, 691, 709 and 716 in Figure 4. Methylation was assayed by sequencing DNA from samples following sodium bisulfite treatment of DNA that converts cytosine to uracil but leaves methyl-cytosine unchanged. Bases that are methylated are coded black, unmethylated bases are coded darker gray, and samples with both methylated and unmethylated bases are coded lighter gray. Samples analyzed included 9 colon cancer cell lines that do not show SLC5A8 transcript expression, 3 colon cancer cell lines that express SLC5A8 transcript, and 6 normal colon tissues. Clearly most colon cancers show substantially more methylation across this region than do normal colon tissues.

To detect the methylation associated with colon cancer a set of methylation specific PCR primers were fashioned. DNA from the assayed tissues was first treated with sodium bisulfite to convert cytosine to uracil, leaving methyl-cytosine unchanged. PCR primers were designed specific for the bisulfite converted sequences arising from methylated or unmethylated templates from the anti-sense strand of the target region (note that after bisulfite conversion the sense and anti-sense strands are no longer complementary to one another).

Figure 8 shows the wild-type sequence of the anti-sense strand of AC063951 between bases 82200-83267. Indicated on this diagram is the position of the MS-PCR1 primers (AS-meth) and the UMS-PCR1 primers (AS-unmethy). The methyl specific MS-PCR1 primers amplify a CpG sites numbered 6, 7, 8 and 15, 16, 17, 18 respectively in Figure 7. The UMS-PCR1 primers interrogate CpG sites 7, 8 and 15, 16, 17, 18 respectively.

Figure 9 shows a blow up of the region and the sequences of the antisense strand that are amplified by the methyl-specific and unmethyl-specific PCR primers.

Figure 10 corresponds to Figure 8, but does not show the wild-type sequence of the anti-sense strand, but the bisulfite converted sequence of a uniformly methylated antisense strand. Indicated again are the position of the methylation specific PCR primers for the MS-PCR1

Figure 11 also corresponds to Figure 8, but does not shows the wild-type sequence of the antisense strand, but the bisulfite converted sequence of a uniformly unmethylated antisense strand. Indicated are the position of the unmethylation specific PCR primers for the UMS-PCR1 assay.

Figure 12 discloses the bisulfite converted sequence of the unmethylated sense strand of nucleotides 82200-83267 of AC063951, renumbered such that basepair 82200 is designated as nucleotide 1.

Figure 13 similarly discloses the bisulfite converted sequence of a uniformly methylated sense strand of nucleotides 82200-83267. To one skilled in the art these disclosures would permit design of methylation specific PCR primers directed against the bisulfite converted sequences of either the sense or antisense strands of the region 82200-83267 demonstrated herein as enabling the detection of human colon cancers.

Figure 14 shows the tabular results of MS-PCR1 assay performed on 31 colon cancer cell lines that do or do not express the SLC5A8 transcript. 70% of cell lines that do not express SLC5A8 score as methylated in the MS-PCR1 assay. No methylation is detected in any cell line that expresses SLC5A8 (100% specificity for prediction of SLC5A8 expression).

Figure 15 shows the tabular results of MS-PCR1 assay performed on 63 matched sets of primary colon cancer tumor tissue and accompanying normal colon tissue. The assay detects 59% of all colon cancers. No methylation was detected in any of 26 normal tissues from patients with unmethylated colon cancers. 3 individuals with MS-PCR1 positive methylation assays in their cancers also showed positivity in their normal colon tissue. It is likely that this represents detection of microscopic contamination of these tissues by tumor cells.

To further test that assertion, Figure 16 gives the results of testing 12 normal colon tissues from individuals without colon cancer. None of the tissues test positive in the MS-PCR1 test. We therefore estimate the sensitivity of MS-PCR1 for detecting colon cancer at 59% and the specificity at 100%.

Figure 17 gives the tabular results of the MS-PCR1 assay of 28 premalignant colon adenomas, 68% of which are detected.

Figure 19 shows RT-PCR detection of the SLC5A8 transcript in normal colon and in a minority subset of colon cancer cell lines, but also demonstrates that 23 of 31 colon cancer cell lines do not express SLC5A8.

Figure 20 shows RT-PCR detection of SLC5A8 transcript in colon cancer cell lines that have been treated with the DNA-demethylating agent 5-azacytidine. 5-azacytidine reactivates expression of the SLC5A8 gene in 6 of 8 colon cancer cell lines, strongly consistent with DNA methylation as the cause of silencing of the SLC5A8 transcript.

Figure 21 demonstrates detection of methylation of the SLC5A8 locus by showing resistance of the locus to HpaII digestion. The 4 HpaII assay (as described in the invention disclosure) is based on PCR amplification of a portion of the SLC5A8 locus. Lanes labeled U show control amplification of undigested SLC5A8 DNA. Lanes labeled M show amplification of DNA that has first been cut with the restriction enzyme Msp1. Msp1 digestion of the DNA eliminates the ability to amplify the locus. Lanes labeled H show amplification of DNA that has first been cut with the restriction enzyme HpaII. HpaII cuts the same sequence as Msp1, but unlike Msp1, HpaII is blocked by DNA methylation. The presence of amplified HpaII cut DNA indicates methylation of the DNA in cell lines V5, V6, RKO, V432, HCT116, V5, V6, V489.

Figure 22 demonstrates detection of SLC5A8 DNA methylation in primary colon cancer tumors but not in matched normal tissue from the same patients. Samples labeled T represent colon cancer tumor tissue; whereas samples labeled N represent the matched normal tissue. Detecting a PCR amplified band after HpaII digestion (lanes labeled H) indicates methylation of the SLC5A8 locus. Methylation of tumor but not normal tissue is seen in samples 529, 365, and 23-21.

Example 2:

A. Identification of the SLC5A8 gene.

Methylation events in genomic DNA from 12 colon cancer cell lines were profiled by restriction landmark genomic scanning. Out of 1,231 unselected CpG islands visualized, spot 3D41 was detected as absent and presumptively methylated in 11 of the 12 colon cancer cell lines. A 510 base pair genomic fragment surrounding the 3D41 site was cloned and shown to correspond to genomic sequence on human chromosome 12q22-23. RNA from normal human colon mucosa was used for connection RT-PCR that linked together over 10 EST sequences

mapping to this genomic region. New sequence was generated both by sequencing of these RT-PCR amplified products, as well as by sequencing image clones corresponding to these ESTs (Figure 28). This established that the 3D41 site was included within a new transcript encoded by a novel gene (Figure 23B). This gene, located on chromosome 12q22-23 gene, is comprised of 15 exons, with the site from RLGS located in exon 1 (Figure 23A). The newly identified transcript includes an in frame TAA stop codon 5' to the presumptive ATG start codon, which additionally is embedded within a GCCATGG sequence that conforms to the standard for a good Kozak sequence. BLAST alignment of the predicted protein product of this novel transcript showed the most closely related proteins to be the human sodium iodide symporter-SLC5A5 (46% homology) and the human sodium-dependent multivitamin transporter- SLC5A6 (43% homology), both of which belong to the solute carrier 5 family (SLC5) of sodium coupled transporters (Figure 29). Moreover, analysis of the predicted novel protein by the TMHMM prediction program (http://www.cbs.dtu.dk/services/TMHMM/) identified 13 transmembrane fragments, which is consistent with structural features of the sodium iodide symporter. Thus structurally, this new transcript encodes a novel member of the SLC5 sodium solute symporter family (SSF) family, and HUGO assigned the encoded protein the name of SLC5A8. A mouse protein of unknown function shows 77% identity to SLC5A8, and is likely the mouse homologs of the human protein (Figure 29). RT-PCR confirmed SLC5A8 transcript was expressed by normal colon mucosa, as well as by kidney, lung, esophagus, small bowel, stomach, thyroid, and uterus, with greatest expression seen in kidney.

B. SLC5A8 is frequently silenced and methylated in colon cancer cell lines.

RT-PCR was used to further characterize SLC5A8 expression in normal colon mucosa compared to a collection of 31 colon cancer cell lines. Whereas the SLC5A8 transcript was well expressed in normal colon, it proved absent in 23 of the 31 colon cancer cell lines (Figure 24A). The methylation of SLC5A8 exon 1 detected by RLGS suggested the hypothesis that aberrant methylation might be the mechanism for silencing of SLC5A8 expression. Consistent with this hypothesis, treatment of SLC5A8 silenced cell lines with the demethylating agent 5-azacytidine reactivated SLC5A8 expression in 6 of 8 colon cancer cell lines tested (Figure 24B and data not shown). Sequencing of the SLC5A8 transcript in the 8 colon cancer cell lines in which it was expressed showed only wild-type sequence with no mutations. Thus methylation, but not mutation, appeared to be the putative mechanism for inactivating SLC5A8 in colon cancer.

To identify target sequences for aberrant SLC5A8 methylation in colon cancer, we investigated a dense CpG island (G+C%=70%, CG/GC=0.9) located in SLC5A8 Exon 1, and surrounding the 3D41 site. This region covered 573 base pairs and included 62 CpG dinucleotides (Figure 30A). In contrast, the region immediately 5' of exon 1 showed only a 46% G+C content. We used sodium bisulfite treatment of genomic DNA to convert unmethylated cytosines to uracil; while leaving methylated cytosines unchanged (Herman and Baylin, 1998, Current Protocols in Human Genetics, N. E. A. Dracopoli, ed., John Wiley & Sons, 2:10.6.1-10.6.10). Sequencing of PCR amplified bisulfite converted SLC5A8 exon 1 genomic DNA was then used to determine the methylation status of each of the 62 target cytosines within the CpG island domain. Comparing the findings in nine SLC5A8-silenced cell lines versus those in three SLC5A8-expressing cell lines and in six samples of SLC5A8 expressing normal colon mucosa defined a 182 bp subregion. In the nine SLC5A8-silenced cell lines this subregion demonstrated uniform methylation of all CpG cytosines; whereas, these cytosines were uniformly unmethylated in the three SLC5A8 expressing cell lines and six normal colon mucosa samples (Figure 30B). Primers for assay of this subregion by methylation specific PCR (MS-PCR) were designed, such that following bisulfite conversion amplification products would selectively be derived from either methylated (M) or unmethylated (U) genomic templates (Herman and Baylin, 1998, Current Protocols in Human Genetics, N. E. A. Dracopoli, ed., John Wiley & Sons, 2:10.6.1-10.6.10). MS-PCR assay of 31 total colon cancer cell lines demonstrated SLC5A8 exon 1 methylation was present in 16 cases (52%), and in each of these methylated cell lines, no SLC5A8 transcript was detectable (Figure 24C). In contrast, in each of the 8 SLC5A8 expressing cell lines MS-PCR assayed exon 1 as unmethylated (Figure 24D). In 7 remaining instances, SLC5A8 expression was absent, but aberrant methylation was not detected as the reason. Moreover, in the case of two of the SLC5A8-methylated cell lines (V425 and V670), DNA from antecedent tumor and matched patient normal tissue was also available. In each of these cases, MS-PCR confirmed that SLC5A8 methylation was present in the primary tumor tissues, but was absent in the matched normal tissues (Figure 24F). Thus the SLC5A8 methylation and silencing detected in colon cancer cell lines reflects somatic aberrations present in primary colon cancer tissues. We note that the finding of gene silencing associated with aberrant methylation in a first exon region corresponding to 5' untranslated sequences has existing precedent at other loci (Attwood et al, 2002, Cell Mol Life Sci 59: 241-257; Jones, P. A. 1999, Trends Genet 15: 34-37).

In previous studies our group has noted that in colon cancers aberrant methylation of hMLH1 and of HLTF commonly silences both maternal and paternal alleles in the same tumor Veigl, et al., 1998, Proc Natl Acad Sci U S A 95:8698-702; Moinova, et al., 2002, Proc Natl Acad Sci U S A 99:4562-7). Consistent with this mechanism, testing of microsatellite markers D12S1041 and D12S1727, that flank *SLC5A8*, showed the presence of two distinguishable parental *SLC5A8* chromosomal regions in 10 of 10 colon cancer cell lines that showed the presence of only methylated *SLC5A8* exon 1.

C. SLC5A8 methylation is commonly present in primary colon cancers and in colon adenomas.

To further establish the frequency of SLC5A8 exon 1 methylation in primary colon cancer tumors, we analyzed by MS-PCR an additional 64 pairs of primary colon cancer tumor tissues as well as their accompanying matched normal colon tissues. SLC5A8 methylation was detected in 38 of 64 (59%) primary colon cancers (Figure 24F and Table 2 below). In 35 of 38 cases (92%) in which colon tumors showed SLC5A8 methylation, this methylation was not detected in the same individuals' normal colon tissues. SLC5A8 exon 1 methylation thus substantially arose in these individuals' cancers as part of and during the neoplastic process. In 3 cases in which SLC5A8 methylation was detected in both an individuals' cancerous and normal colon tissues, these findings likely indicate either the presence of some cancer cells within the grossly normal resected tissue, or the possibility that the cancer arose from a field of SLC5A8 methylated cells. The rarity of detecting SLC5A8 methylation in normal colon tissues is highlighted by noting that no SLC5A8 methylation was detected in any of the 26 normal colon tissues in which the accompanying colon cancer was also unmethylated (Table 2 below), and moreover, that no SLC5A8 methylation was detected in any of 12 additional normal colon tissues from resections done for non-cancer diagnoses.

Table 2. SLC5A8 Methylation in Colon Tumors and Matched Normal Mucosa. Shown is the characterization of 64 pairs of colon cancer tumors and matched normal colon tissues assayed for methylation of SLC5A8 exon 1 by MS-PCR. Indicated are the numbers (and percentages) of tissue pairs with each of the four possible methylation phenotypes.

		NORMAL TISSUE	
		Methylated	Unmethylated

TUMOR TISSUE	Methylated	3 (5%)	35 (54%)
	Unmethylated	0 (0%)	26 (41%)

Among all primary cancers and cell lines analyzed, the finding of SLC5A8 methylation in colon cancer tumors and cell lines was not significantly correlated with either patients' sex (P=0.39) or age (P=0.52), with a median age of 69 in persons with SLC5A8-methylated cancers versus 67 in those with SLC5A8 unmethylated cancers. Moreover, the distribution by tumor stage (Dukes' stage B, C, D primary tumor; or metastatic cancer deposit) was not significantly different between SLC5A8-methylated and nonmethylated colon cancers (P=0.77) (Table 3 below). SLC5A8 methylated and unmethylated cancers also showed no significant difference with respect to site of origin in the rectum, left colon, or right colon (P= 0.47) (Table 4 below).

Table 3. Distribution of SLC5A8 methylation by tumor stage. Shown are numbers (and %) of colon neoplasms (tumor and cell lines) in each category defined by clinical stage and SLC5A8 methylation status.

Tumor Stage	SLC5A8 Methylated	SLC5A8 Unmethylated
Adenoma	17(24%)	12 (23%)
Duke's B	24 (34%)	16 (30%)
Duke's C	15 (21%)	13 (25%)
Duke's D	6 (8%)	5 (9%)
Metastatic lesion	7 (10%)	7 (13%)

Table 4. Distribution of SLC5A8 methylation by tumor site. Shown are numbers (and %) of colon neoplasms (tumor and cell lines) in each category defined by location in the colon and SLC5A8 methylation status.

Tumor site	SLC5A8 Methylated	SLC5A8 Unmethylated

Right colon	12 (23%)	13 (35%)
Left colon	30 (59%)	20 (54%)
Rectal	9 (18%)	4 (11%)

To determine the timing of onset of SLC5A8 silencing during colon carcinogenesis, we additionally analyzed a group of 29 adenomas for SLC5A8 exon 1 methylation. SLC5A8 methylation was detected in 17 of the 29 (59%) adenoma cases. SLC5A8 methylation thus appears to be an early event that is already established in colon neoplasia by the adenoma stage.

D. Quantitative assay of SLC5A8 exon 1 methylation.

To derive a quantitative measure of SLC5A8 methylation, we employed a real time MS-PCR assay whose results were expressed as 1000 times the ratio of methylated SLC5A8 reaction product to a control *MYOD1* reaction product (Usadel, et al., 2002, Cancer Res 62:371-5). In this assay, 0 methylation was detected in the Vaco9 SLC5A8 expressing colon cancer cell line, and a methylation value of 1000 was detected in the SLC5A8 methylated and silenced RKO colon cancer cell line. As shown in Figure 25A, assay for SLC5A8 exon 1 methylation in 11 normal colon mucosal samples derived from non-cancer resections yielded only barely detectable methylation values (mean value= 24; range= 4 –82) and defined an "unmethylated normal range" of values all < 100. Analysis of 29 normal colon samples derived from colon cancer resections gave similarly low values with a mean value =22 and with a single outlier sample (value =159) falling outside the range defined by the non-cancer derived normal tissues. This observation essentially replicated our previous observation of rare faint methylation events detected in some cancer associated normal tissue. In contrast, analysis of colon cancer samples clearly distinguished two populations of tumors. Twelve cancers were deemed unmethylated, as they showed methylation values falling well within the population normal range (mean value =12; range = 0-58) (Figure 25A), and hence were indistinguishable from unmethylated normal tissues. In contrast, 17 cancers with methylation values greater than the normal range comprised a distinct "methylated" group of cancers that was characterized by a mean methylation value of 747 and a range = (121- 2549) (Figure 25A). The mean methylated colon cancer thus displayed 75% the level of methylation as was measured in a pure cell line population of methylated RKO cells. The heterogeneity in measured methylation values among the methylated colon cancers

may in part derive from differences among the tumors in levels of contaminating and infiltrating non-cancer cells. The methylated and unmethylated cancer populations defined by real time MS-PCR respectively corresponded to the tumors classified as unmethylated and methylated in the previous non-quantitated MS-PCR reaction.

E. Detection of SLC5A8 methylation in aberrant crypt foci.

The finding of SLC5A8 methylation in colon adenomas prompted us to consider that SLC5A8 methylation might be an early event in human colon neoplasia. The earliest morphologically identifiable colon neoplasias putatively are aberrant crypt foci (ACF) (Siu et al., 1999, Cancer Res 59: 63-66). These microscopic morphologically aberrant multicrypt structures are recognizable in unembedded colon under low power magnification. Moreover, a subset of ACF lesions demonstrate both histologic dysplasia and mutations of the *APC* tumor suppressor gene (Bird, 1987, Cancer Lett 37:147-51; Pretlow, et al., 1991, Cancer Res 51:1564-7), suggesting that at least some ACF have potential to progress to colon adenomas and cancers. To assess a possible role of SLC5A8 methylation in ACF development, 15 ACF, composed of from 17 to 155 crypts (48±36 crypts, mean ± standard deviation), were dissected from 11 different patients' colons bearing either cancer or adenomas. From these same 11 cases, 24 similarly sized tissue samples were dissected from mucosal regions that appeared normal under low power magnification. Real time MS-PCR analysis of SLC5A8 methylation in the 24 control normal samples gave results similar to those obtained in previous normal mucosal samples, with a mean SLC5A8 methylation value of 12, and with only one of these 24 new samples (methylation value of 117) falling just outside of the previously determined normal limit of 100 (Figure 25B). In contrast, analysis of DNA from the ACF revealed two distinct populations, with 8 of 15 ACF falling within the normal range (mean =34, and range =0-113), and with 7 of 15 ACF samples demonstrating SLC5A8 values that fell well within the range of methylated cancers (mean =355, range =287-420) (Figure 25B). In contrast, none of these 15 aberrant crypt foci demonstrated aberrant methylation of *hMLH1*, which thus likely arises later during colon carcinogenesis. These findings suggest that SLC5A8 methylation is indeed an early aberration that precedes adenoma formation and is detectable in aberrant crypt foci. This finding also further strengthens the model that suggests a subset of aberrant crypt foci are likely to progress to more advanced colonic neoplasms.

F. SLC5A8 methylation as a serologic marker of colon cancer.

SLC5A8 methylation was detected in 59% of our primary colon samples. In these same samples we had previously noted a 44% frequency of methylation of *HLTF*, a SWI/SNF family gene (Moinova et al., 2002, Proc Natl Acad Sci USA 99: 4562-4567), and had also found a 44% frequency of methylation of *p16* (Figure 31) (Herman et al., 1995, Cancer Res 55: 4525-4530; Gonzalez-Zulueta et al., 1995, Cancer Res 55: 4531-4535). These data suggest SLC5A8 methylation might be a high quality marker of colon cancer presence. In this regard, we and others have shown that aberrantly methylated genomic DNA from specific loci can be detected in the serum of some cancer patients (Grady et al., 2001, Cancer Res 61: 900-902; Hibi et al., 1998, Cancer Res 58: 1405-1407; Jeronimo et al., 2001, J Natl Cancer Inst 93: 1747-1752; Usadel et al., 2002, Cancer Res 62: 371-375). Accordingly, we characterized the level of SLC5A8 methylation in ethanol precipitable DNA prepared from the serum of colon cancer patients (Grady et al., 2001, Cancer Res 61: 900-902). SLC5A8 methylation was totally undetectable with a measured value of 0 in DNA extracted from each of 13 serum samples from individuals with colon cancers in which SLC5A8 assayed as unmethylated (Figure 26). In contrast, SLC5A8 methylation was detectable in serum DNA from 4 of 10 patients in which the underlying colon cancer assayed as SLC5A8 methylated (Figure 26). A positive signal for *MYOD1* verified the presence of input DNA into each of these assays. While serologic assays for methylated DNA as a marker of cancer are clearly in the early stages of investigation, we note that a panel of methylated genes that included SLC5A8, HLTF, p16 and hMLH1 provided greater sensitivity than any single locus alone for detecting an aberrant methylation event in our set of 64 primary colon cancers (Figure 31).

G. SLC5A8 suppression of colon cancer colony formation.

The high frequency of SLC5A8 methylation observed in colon cancer suggested that inactivation of this gene might confer a selective advantage. To assay for such an advantage, we examined the effect of SLC5A8 transfection in three colon cancer cell lines (V400, RKO and FET) in which the endogenous SLC5A8 gene was methylated and silenced, as compared with three colon cancer cell lines (V457, V9M and V364) in which the endogenous SLC5A8 gene remained unmethylated and expressed. Reconstitution of SLC5A8 expression in SLC5A8-methylated cells suppressed colony-forming ability by at least 75% in each of the three lines tested ($P<0.01$) (Figure 27B). In contrast, transfection of SLC5A8 did not show significant colony suppression in the any of the three cell lines that already expressed an endogenous SLC5A8 allele (Figure 27A) ($P< 0.01$ for the difference in effect of SLC5A8 transfection in methylated versus unmethylated cell lines). Transient transfection showed that both

SLC5A8-methylated and unmethylated cells were able to express comparable levels of exogenous SLC5A8, as determined by western analysis for a V5 epitope tag attached to the SLC5A8 cDNA. These findings suggest that SLC5A8 methylation and silencing confers a specific growth advantage in the subset of colon cancers in which this locus is inactivated.

Consistent with this interpretation, we found that 4 of 5 of the rare SLC5A8 expressing clones that grew out following transfection of the *SLC5A8* methylated V400 colon cancer cell lines were markedly suppressed in their ability to form xenograft tumors in athymic mice (Figure 32).

H. Discussion.

In this study, we have identified a novel gene, SLC5A8, that we demonstrate is a new candidate colon cancer suppressor gene. We find that SLC5A8 encodes a sodium transporter and is a new member of the sodium solute symporter family (SLC5). SLC5A8 is frequently targeted for methylation and silencing in human colon cancer, with aberrant SLC5A8 exon 1 methylation was detected in 52% of colon cancer cell lines and in 59% of primary colon cancers. All colon cancer cell lines showed that SLC5A8 exon 1 methylation were silenced for SLC5A8 expression, and SLC5A8 expression could be restored by treatment with a demethylating agent 5-azacytidine. We therefore conclude that epigenetic gene silencing, which is reflected by aberrant SLC5A8 methylation represents the principal mechanism for inactivating this gene in colon cancer. Moreover, our finding that exogenous SLC5A8 specifically suppresses colony forming activity in colon cells that have inactivated this allele supports the hypothesis that SLC5A8 inactivation confers a selectable advantage in neoplastic colon epithelial cells. Colon cells that retain SLC5A8 are insensitive to the introduction of an exogenous allele, and presumably bear a mutation elsewhere that renders them tolerant to continued SLC5A8 expression. Also supporting that SLC5A8 methylation is a pathogenetic event in colon neoplasia is our finding that SLC5A8 methylation is a highly early event that is detectable in 47% of aberrant crypt foci, which are the earliest detectable morphologic abnormality of the colon epithelium.

SLC5A8 methylation may also play an etiologic role in malignancies additional to colon cancer. In earlier studies, we note that SLC5A8 methylation is present in a subset of cancers of the breast and stomach cancers (Table 5 below).

Table 5. SLC5A8 methylation in additional cancers. Shows are the results of MS-PCR assay for SLC5A8 exon 1 methylation in primary human tumors. In each case, paired normal tissue assayed as unmethylated.

Cancer Types	Breast	Stomach	Kidney
SLC5A8 methylated	4	4	0
SLC5A8 unmethylated	16	2	7

Both molecular homology and functional data suggest that SLC5A8 functions as a sodium solute symporter. There are 109 currently known members of the sodium solute symporter family which functions to co-transport sodium coupled to solutes as diverse as iodine (NIS/SLC5A5), glucose (SGLT1/SLC5A1; SGLT2/SLC5A2), inositol (SMIT/SLC5A3), and water soluble vitamins (SMVT/SLC5A6) (Smanik et al., 1996, Biochem Biophys Res Commun 226: 339-345; Prasad et al., 1998, J Biol Chem 273: 7501-7506; Wright et al., 1994, J Exp Biol 196: 197-212). Elucidating the putative solute cotransported by SLC5A8 may provide future insight both into the mechanism of SLC5A8 growth suppression, as well as leads for potential development of novel agents useful for colon neoplasia prevention and treatment.

Materials And Methods

Sequences. Human *SLC5A8* mRNA and gene sequence accession numbers as deposited by our group are AF53621 and AF536217. The *SLC5A8* murine homolog is accession number is BC017691. Contemporaneously with our Genbank entry, *SLC5A8* mRNA sequence was also independently deposited under accession number AY081220 (Rodriguez et al., 2002, J Clin Endocrinol Metab. 87:3500-3).

Restriction Landmark Genomic Scanning (RLGS). RLGS was performed as previously described (Costello et al., 2000, Nat Genet 24: 132-138).

Amplification and Sequencing of *SLC5A8*. The primers used for RT-PCR assay of a *SLC5A8* fragment are 5'-TCCGAGGTCTACCGTTTTG-3', and 5'-GGGCA GGGGC ATAAA TAAC-3'. The PCR parameters were 35 cycles of 95°C (45s), 54°C (45s), 72°C (60s), 72°C

(10min), and 4°C to cool. The full length *SLC5A8* ORF was amplified using primers: 5'-TCCGGGATAAGAAGTGCG-3' and 5'-TAGTATCAGAGCAGCTTCACAAAC-3'. GC-rich cDNA polymerase kit (Clonetech) was used and PCR parameters were 35 cycles of 95°C (45s), 62°C (45s), 72°C (90s), 72°C (10min), and 4°C to cool. Sequencing primers were: 5'-TTTGT GGTGGTCA TCAGCG-3', 5'-GGGCAGGGGCATAAATAAC-3', 5'-AGGCTGTG GTGATGCAAGGT-3', 5'-TTAATGCCTTAGCAGCAG-3', and 5'-CCTCCACTT CCTGAGAGAAC-3'.

Constructs. To construct the V5 tagged SLC5A8 expression vector, the following PCR primers were used: 5'-TCCGGGATAAGAAGTGCG-3' and 5'-TCTAGTATCA GAGCAGCTACACAA-3'. The PCR conditions were the same as employed for amplification of the full length ORF. PCR products were cloned into pcDNA3.1/V5-His-TOPO vector (Invitrogen).

Serum DNA purification. Blood was drawn into red/grey vacutainer collection tubes and allowed to clot for 2 hours. It was then spun in a clinical table top centrifuge for 15 min at 3000 rpm at room temperature. Serum was collected using a sterile pipette, divided into 1 ml aliquots, and stored at -80°C. Serum DNA from patients was purified as described previously (Grady et al., 2001, Cancer Res 61:900-902).

Western Analysis. Approximately 10^7 cells were lysed in cell lysis buffer [50 mM Tris.HCl (pH 7.4)/1 mM EGTA/1% Nonidet P-40/0.25% sodium deoxycholate/150 mM NaCl]. Equal amounts of protein were subjected to SDS polyacrylamide gel electrophoresis and then transferred to a PVDF nylon membrane (Millipore), which was probed with 1:200 dilution of mouse anti-V5 monoclonal antibody (Invitrogen). Immune complexes were visualized with ECL+Plus Western blotting detection kit (Amersham) after incubation with horseradish peroxidase-coupled secondary antibody (Santa Cruz).

Sodium Bisulfite Treatment: Flanking PCR and MS-PCR. Sodium bisulfite treatment to convert unmethylated cytosine to thymidine was performed similarly as described (Grady et al., 2001, Cancer Res 61:900-902). Primers that flank the SLC5A8 exon 1 CpG island are 5'-CGTGAA GGTAAA GATGTT AAAAATG-3' and 5'-ACAACT AAAAAC TCCAAT TCTCATC-3'. PCR were carried out by using a hot start at 95°C (7 min) and following cycling parameters: 35 cycles of 95°C (45s), 56°C (45s), 72°C (45s), 72°C (10 min), and 4°C to cool. Primers to amplify the methylated allele are AS-meth-442-459s: 5'-TCGAAC GTATTT

CGAGGC-3' and AS-meth-550as: 5'-ACAACG AATCGA TTTTCCG-3'. PCR parameters are 31 cycles of 95°C (45s), 56°C (45s), 72°C (45s), 72°C (10 min), and 4°C to cool. Primers to amplify the unmethylated allele are AS-unmeth-442s: 5'-TTGAAT GTATTT TGAGGTG-3' and AS-unmeth-542as: 5'-TCAATT TTCCAA AATCCC-3'. PCR parameters are 31 cycles of 95°C (45s), 46°C (45s), 72°C (45s), 72°C (10min), and 4°C to cool.

Methylation-Specific Real-time PCR. The same MS-PCR primers as above (As-meth-442-459s and As-meth-550as), were first used to amplify a bisulfite converted methylated SLC5A8 exon 1 template. A fluorogenic hybridization probe was designed using sequences specific for the sodium bisulfite converted SLC5A8 methylated template. The sequence was the following: 5'-6FAM-CAACGACGAAT ACAAAAACG ACTACCAAC-BHQ-2-3'. Bisulfite converted sequences from the *MYOD1* gene were used as an internal reference as described by (Usadel et al., 2002, Cancer Res 62: 371-375). Primers and probes for *MYOD1* were: forward primer: 5'-CCAACTCCA AATCCCCTC TCTAT-3'; reverse primer: 5'-TGATTAATT TA GATTGGGTTT AGAGAAGGA-3'; and probe: 5'-6FAM-TCCCTTCCT ATTCCTAAA TCCAAC CTAAATACCTCC-BH-2-3'. All the above primers and probes were synthesized by Integrated DNA Technologies, Inc. For the gene of interest, SLC5A8, the reaction mix contained 600 nM primer, 200 nM probe, 5.5 mM-Mg^{2+}, 1X Supermix from Bio-Rad. The total volume was 25 µl. For the *MYOD1* gene, the reaction mix contained 400 nM primer, 200 nM probe, 3 mM-Mg^{2+}, 1X Supermix from Bio-Rad. The total volume was also 25 µl. Thermal cycling was initiated with 50°C for 2 min, then 95°C for 10 min, followed by 55 cycles of 95°C for 15 sec and 60°C for 1 min. PCR was performed in separate wells for each probe/primer set. Each plate contained multiple positive controls, negative controls and water blanks. Colon cancer cell line RKO was used for a positive control, and V9M as a negative control. Serial dilutions of RKO DNA were used to create a standard curve. SLC5A8 methylation was determined as the ratio of SLC5A8:MYOD1=2 exp- (CT_{SLC5A8}- CT_{MYOD1}).

Aberrant Crypt Foci. Aberrant crypt foci (ACF) (Bird, 1987, Cancer Lett 37: 147-151; Pretlow et al., 1991, Cancer Res 51: 1564-1567; Siu et al., 1999, Cancer Res 59: 63-66) were isolated from grossly normal human colonic mucosa according to the method of Bird et al. (Bird et al., 1997, Cancer Lett 116: 15-19). Strips of human colonic mucosa, stored over liquid nitrogen, were thawed rapidly in 1% paraformaldehyde and fixed flat in 70% ethanol for 30 min at 4°C (Bird et al., 1997, Cancer Lett 116: 15-19). The colonic strips were stained for 2 min in 0.2% methylene blue (Chroma-Gesellschaft Schmid & Co, distributed by Roboz Surgical Co, Washington, DC) in 0.1 M sodium phosphate buffer (pH 7.4), rinsed in 1%

paraformaldehyde for 15 min, transferred mucosal side up to a glass slide and viewed at 30X magnification under a dissecting microscope. The ACF were teased from the mucosa with microdissection forceps (FWR #55 Dumont Bio Inox Forceps, 0.05 x 0.02 mm tips), placed in microfuge tubes, and stored over liquid nitrogen. The control for each ACF was a similar number of microscopically normal crypts teased from the same mucosa.

Cell Culture and Clonogenic Assays. Vaco cell lines were cultured as previously described (Veigl et al., 1998, Proc Natl Acad Sci U S A 95: 8698-8702; Markowitz et al., 1995, Science 268: 1336-1338; Willson et al., 1987, Cancer Res 47: 2704-2713). FET and RKO were the kind gift of Dr. M. Brattain (Roswell Cancer Institute, Buffalo, NY). Colony formation assays were performed as described (Moinova et al., 2002, Proc Natl Acad Sci USA 99: 4562-4567). Briefly, colon cancer cells were plated on a rat tail collagen matrix (Willson et al., 1987, Cancer Res 47: 2704-2713) (which was found necessary for proper membrane localization of SLC5A8 protein). Cells were then transfected with either a SLC5A8 expression vector or a control empty vector, and the number of stable colonies arising after selection in G418 was respectively counted.

5-Azacytidine Treatment. The treatment was performed as described previously (Veigl et al., 1998, Proc Natl Acad Sci U S A 95: 8698-8702). Briefly, cells were treated for 24 h on day 2 and day 5 with 5-azacytidine (Sigma) at 1.5 μg/ml. The medium was changed 24 h after addition of the 5-azacytidine (i.e., on day 3 and day 6).

Statistical Methods. Association of SLC5A8 methylation with sex was analyzed by using two-tailed Fishers' exact tests. Association of SLC5A8 methylation status with tumor site or stage was analyzed by using Pearson's χ^2 statistics. Comparisons of age distributions based on SLC5A8 methylation were done by using Wilcoxon nonparametric tests. Comparisons of colony counts after transfection with different vectors were done by *t* tests and linear models.

Hap2 site assays. (1) For 4 Hpa2 site assays, the following primers were used: 5'-CCAGCGAAGGCGTAGTAGAT-3' (3D41-Hpa2-190R) and 5'-GGCTCCAGTTCTCA TCTGCT-3' (3D41-Hpa2-633F). The Advantage-GC-genomic DNA polymerase kit was used. Thermal cycling was performed at 95°C for 1 min, 95°C for 45 sec, 63°C for 45, 72°C for 90 sec, then followed by 26 cycles, and finally 72°C for 5 min. (2) For 6 Hpa2 site assays, the following primers were used: 5'-CCAGCGAAGGCGTAGTAGAT-3' (3D41-Hpa2-190R) and 5'-GGCAGTCTAAAAACTCCAGGC-3' (3D41-Hpa2-82430F). The Advantage-GC-genomic

DNA polymerase kit was used. Thermal cycling was performed at 95°C for 7 min, 95°C for 45 sec, 64°C for 45, 72°C for 90 sec, then followed by 29 cycles, and finally 72°C for 5 min. In both assays, aberrant methylation of colon cancer cells is indicated by recovery of a PCR product from DNA that has been digested with the restriction enzyme Hpa2.

Incorporation by Reference

All publications and patents mentioned herein are hereby incorporated by reference in their entirety as if each individual publication or patent was specifically and individually indicated to be incorporated by reference. In case of conflict, the present application, including any definitions herein, will control.

Equivalents

While specific embodiments of the subject invention have been discussed, the above specification is illustrative and not restrictive. Many variations of the invention will become apparent to those skilled in the art upon review of this specification and the claims below. The full scope of the invention should be determined by reference to the claims, along with their full scope of equivalents, and the specification, along with such variations.

We claim:

1. An isolated polypeptide comprising an amino acid sequence selected from the group consisting of:
a) an amino acid sequence at least 95% identical to SEQ ID NO: 1; and
b) an amino acid sequence encoded by a nucleic acid that hybridizes under high stringency conditions to a nucleic acid of any one of SEQ ID NOs: 3 or 4,
wherein said polypeptide is a cell surface protein.

2. The isolated polypeptide of claim 1, wherein the polypeptide comprises a transmembrane domain as set forth in any one of SEQ ID NOs: 19-31.

3 An isolated antibody, or fragment thereof, which is specifically immunoreactive with an epitope of an amino acid sequence as set forth in SEQ ID NO: 1.

4. The antibody of claim 3, wherein said antibody is selected from the group consisting of: a polyclonal antibody, a monoclonal antibody, an Fab fragment and a single chain antibody.

5. The antibody of claim 3, wherein said antibody is labeled with a detectable label.

6. An isolated nucleic acid selected from the group consisting of:
a) a nucleic acid comprising the nucleotide sequence of SEQ ID NO: 2, or a complement thereof;
b) a nucleic acid molecule that encodes a polypeptide comprising the amino acid sequence at least 95% identical to the amino acid sequence of SEQ ID NO: 7; and
c) a nucleic acid molecule that hybridizes under stringent conditions to SEQ ID NO: 2.

7. The nucleic acid of claim 6, further comprising a vector nucleic acid sequence.

8. A host cell which contains the nucleic acid of claim 6.

9. A method for producing the polypeptide of claim 1, comprising culturing the host cell of claim 5 under conditions in which the nucleic acid molecule is expressed.

10. A method for detecting the presence of the polypeptide of claim 1 in a sample, comprising: a) contacting the sample with an antibody which selectively binds to the polypeptide of claim 1; and b) determining whether the antibody binds to the polypeptide in the sample.

11. A kit for detecting a human SCL5A8 polypeptide comprising: (i) an antibody of claim 3; and (ii) a detectable label for detecting said antibody.

12. A method for detecting the presence of the nucleic acid of claim 6 in a sample, comprising:
a) contacting the sample with the probe or primer of claim 6; and
b) determining whether the probe or primer binds to a nucleic acid in the sample.

13. A kit comprising the probe or primer of claim 6 and instructions for use.

14. A method for identifying a compound which binds to the polypeptide of claim 1, comprising:
a) contacting the polypeptide, or a cell expressing the polypeptide of claim 1, with a test compound; and
b) determining whether the polypeptide binds to the test compound.

15. A method for modulating the activity of the polypeptide of claim 1, comprising contacting the polypeptide or a cell expressing the polypeptide of claim 1 with a compound which binds to the polypeptide in a sufficient concentration to modulate the activity of the polypeptide.

16. A method of inhibiting aberrant activity of a SLC5A8-expressing cell, comprising contacting the cell with a compound that modulates the activity or expression of the polypeptide of claim 1, in an amount which is effective to reduce or inhibit the aberrant activity of the cell.

17. The method of any of claims 14-16, wherein the compound is selected from the group consisting of a peptide, a phosphopeptide, a small organic molecule, an antibody, and a peptidomimetic.

18. The method of any of claims 14-17, wherein the cell is found in the colon, kidney, lung, esophagus, small bowel, stomach, thyroid, uterus, and breast.

19. A method of treating or preventing a disorder characterized by aberrant activity of a SLC5A8-expressing cell, in a subject, comprising administering to the subject an effective amount of a compound that modulates the activity or expression of the polypeptide of claim 1, such that the aberrant activity of the SLC5A8-expressing cell is reduced or inhibited.

20. A transgenic mouse having germline and somatic cells comprising a chromosomally incorporated transgene that disrupts the genomic SLC5A8 gene and inhibits expression of said gene, wherein said disruption comprises insertion of a selectable marker sequence resulting in said transgenic mouse exhibiting increased susceptibility to the formation of tumors as compared to the wildtype mouse.

21. The transgenic mouse of claim 20, wherein said mouse is homozygous for said disruption.

22. The transgenic mouse of claim 20, wherein said mouse is heterozygous for said disruption.

23. A transgenic mouse having germline and somatic cells in which at least one allele of a genomic SLC5A8 gene is disrupted by a chromosomally incorporated transgene, which transgene inhibits the expression of said genomic SLC5A8 gene, wherein (i) said genomic SLC5A8 gene encodes a SLC5A8 protein; and (ii) said disruption comprises insertion of a selectable marker sequence, which replaces all or a portion of the genomic SLC5A8 gene or is inserted into the coding sequence of said genomic SLC5A8 gene; and (iii) said transgenic mouse has increased susceptibility to the development of neoplasms.

24. Isolated mammalian cells comprising a diploid genome including a chromosomally incorporated transgene, which transgene disrupts the genomic SLC5A8 gene and inhibits expression of said gene.

cells of claim 24, which cells are mouse cells.

26. A method for generating a mouse and mouse embryonic stem cells having a functionally disrupted endogenous SLC5A8 gene, comprising the steps of:
(i) constructing a transgene construct including (a) a recombination region having all or a portion of the endogenous SLC5A8 gene, which recombination region directs recombination of the transgene with the endogenous SLC5A8 gene; and (b) a marker sequence which provides a detectable signal for identifying the presence of the transgene in a cell;
(ii) transferring the transgene into embryonic stem cells of a mouse;
(iii) selecting embryonic stem cells having a correctly targeted homologous recombination between the transgene and the SLC5A8 gene;
(iv) transferring said cells identified in step (iii) into a mouse blastocyst and implanting the resulting chimeric blastocyst into a female mouse; and
(v) selecting offspring harboring an endogenous SLC5A8 gene allele comprising the correctly targeted recombination.

27. A method of evaluating the carcinogenic potential of an agent comprising: (i) contacting the transgenic mouse of claim 20 with a test agent; and (ii) comparing the number of transformed cells in a sample from the treated mouse with the number of transformed cells in a sample from an untreated transgenic mouse or transgenic mouse treated with a control agent, wherein the difference in the number of transformed cells in the treated mouse, relative to the number of transformed cells in the absence of treatment or treatment with a control agent, indicates the carcinogenic potential of the test compound.

28. A method of evaluating an anti-proliferative activity of a test compound, comprising:
(i) providing a transgenic mouse of claim 20 having germline and somatic cells in which the expression of the SLC5A8 gene is inhibited by said chromosomally incorporated transgene, or a sample of cells derived therefrom;
(ii) contacting the transgenic mouse or the sample of cells with a test agent; and
(iii) determining the number of transformed cells in a specimen from the transgenic mouse or in the sample of cells,
wherein a statistically significant decrease in the number of transformed cells, relative to the number of transformed cells in the absence of the test agent, indicates the test compound is a potential anti-proliferative agent.

29. A method for detecting differential methylation patterns in a SLC5A8 nucleotide sequence, comprising:

a) obtaining a sample from a patient;

b) assaying said sample for the presence of methylation within a nucleotide sequence as set forth in any one of SEQ ID NOs: 12-13 or fragments thereof;

c) obtaining a sample from a healthy subject;

d) assaying for the presence of methylation in a nucleotide sequence as set forth in any one of SEQ ID NOs: 12-13 or fragments thereof; and

e) comparing the methylation patterns in the sample from the patient to the methylation patterns in the normal sample.

30. A method for detecting a SLC5A8-associated cancer, comprising:

a) obtaining a sample from a patient; and

b) assaying said sample for the presence of methylation within a nucleotide sequence as set forth in any one of SEQ ID NOs: 12-13 or fragments thereof;

wherein methylation of said nucleotide sequence is indicative of a SLC5A8-associated cancer.

31. The method of any one of claims 29 and 30, wherein the sample is a bodily fluid selected from the group consisting of blood, serum, plasma, a blood-derived fraction, stool, urine, and a colonic effluent.

32. The method of claim 31, wherein the bodily fluid is obtained from a subject suspected of having or is known to have a SLC5A8-associated cancer.

33. The method of claim 32, wherein said SLC5A8-associated cancer is selected from the group consisting of: colon cancer, breast cancer, thyroid cancer, and stomach cancer.

34. The method of any one of claims 29 and 30, comprising assaying for the presence of methylation within the SLC5A8 sequence as set forth in SEQ ID NO: 14.

35. The method of any of claims 29-34, wherein the assay is methylation-specific PCR.

36. The method of claim 35, comprising:

a) treating DNA from the sample with a compound that converts non-methylated cytosine bases in the DNA to a different base;

b) amplifying a region of the compound converted SLC5A8 nucleotide sequence with a forward primer and a reverse primer; and

c) analyzing the methylation patterns of said SLC5A8 nucleotide sequences.

37. The method of claim 35, comprising:

a) treating DNA from the sample with a compound that converts non-methylated cytosine bases in the DNA to a different base;

b) amplifying a region of the compound converted SLC5A8 nucleotide sequence with a forward primer and a reverse primer; and

c) detecting the presence and/or amount of the amplified product.

38. The method of claim 35, wherein the forward primers are selected from SEQ ID NOs: 8 and 10.

39. The method of claim 35, wherein the reverse primers are selected from SEQ ID NOs: 9 and 11.

40. The method of claim 35, wherein the compound used to treat DNA is a bisulfite compound.

41. The method of any of claims 29-34, wherein the assay comprises using a methylation-specific restriction enzyme.

42. The method of claim 41, wherein said methylation-specific restriction enzyme is selected from HpaII, SmaI, SacII, EagI, MspI, BstUI, and BssHII.

43. The method of claim 41, further comprising a pair of primers selected from SEQ ID NOs: 5-7.

44. A method for detecting a SLC5A8-associated cancer in a subject, comprising detecting SLC5A8 protein or nucleic acid expression in a sample from the subject.

45. The method of claim 44, wherein the sample is a bodily fluid selected from the group consisting of blood, serum, plasma, a blood-derived fraction, stool, urine, and a colonic effluent.

46. The method of claim 45, wherein the bodily fluid is from a subject suspected of having or known to have a SLC5A8-associated cancer.

47. The method of claim 46, wherein the SLC5A8-associated cancer is selected from the group consisting of: colon cancer, breast cancer, thyroid cancer, and stomach cancer.

48. The method of claim 44, wherein the SLC5A8 protein is detected by immunoassays.

49. A method for identifying an agent which enhances SLC5A8 protein or nucleic acid expression in a diseased cell associated with SLC5A8 gene silencing, comprising:

a) contacting the cell with a sufficient amount of the agent under suitable conditions;

b) quantitatively determining the amount of SLC5A8 protein or nucleic acid; and

c) comparing the amount of SLC5A8 protein or nucleic acid with the amount of SLC5A8 protein or nucleic acid in the absence of the agent,

wherein a greater amount of SLC5A8 protein or nucleic acid in the presence of the agent than in the absence of the agent indicates that the agent enhances SLC5A8 protein or nucleic acid expression.

50. The method of claim 49, wherein said SLC5A8 gene silencing is due to differential methylation of a SLC5A8 nucleotide sequence.

51. The method of claim 50, wherein differential methylation occurs within a SLC5A8 nucleotide sequence set forth in any one of SEQ ID NOs: 12-13 or fragments thereof.

52. The method of claim 49, wherein the diseased cell is from a subject having colon neoplasia.

53. A method for monitoring over time a SLC5A8-associated cancer comprising:

a) detecting the methylation status of a SLC5A8 nucleotide sequence in a sample from the subject for a first time; and

b) detecting the methylation status of the SLC5A8 nucleotide sequence in a sample from the same subject at a later time;

wherein absence of methylation in the SLC5A8 nucleotide sequence taken at a later time and the presence of methylation in the SLC5A8 nucleotide sequence taken at the first time is indicative of cancer regression;

wherein presence of methylation in the SLC5A8 nucleotide sequence taken at a later time and the absence of methylation in the SLC5A8 nucleotide sequence taken at the first time is indicative of cancer progression.

54. The method of claim 53, wherein the sample is a bodily fluid selected from the group consisting of blood, serum, plasma, a blood-derived fraction, stool, urine, and a colonic effluent.

55. The method of claim 53, wherein the SLC5A8-associated cancer is selected from the group consisting of: colon cancer, breast cancer, thyroid cancer, and stomach cancer.

56. A method for treating a SLC5A8-associated proliferative disease in a subject, comprising administering to the subject a sufficient amount of a compound, wherein the compound modulates the SLC5A8 protein or nucleic acid expression.

57. The method of claim 56, wherein the disease is associated with methylation of a SLC5A8 nucleic acid sequence, and the compound induces SLC5A8 expression.

58. The method of claim 57, the compound is a demethylation agent selected from 5-azacytidine and 5-deoxy-azacytidine.

59. The method of claim 56, wherein the SLC5A8-associated proliferative disease is selected from the group consisting of: thyroid nodular hyperplasia, thyroid adenoma, thyroid cancer, colon neoplasia, breast cancer, and stomach cancer.

60. A method for treating a SLC5A8-associated cancer in a subject, comprising administering to the subject a vector containing a SLC5A8 nucleic acid which is operably linked to a heterologous promoter.

61. The method of claim 60, wherein the SLC5A8 nucleic acid encodes a polypeptide at least 90% identical to SEQ ID NO: 1.

62. The method of claim 60, wherein the cancer is a colon neoplasia.

63. A bisulfite-converted methylated SLC5A8 nucleotide sequence selected from the group consisting of:

a) a nucleotide sequence of any one of SEQ ID NOs: 15-18 or a fragment thereof;

b) a complement of any one of SEQ ID NOs: 15-18; and

c) a nucleotide sequence that hybridizes under stringent conditions to a nucleotide sequence of any one of SEQ ID NOs: 15-18.

64. Oligonucleotide primers for detecting methylation of a SLC5A8 nucleotide sequence, selected from SEQ ID NOs: 5-11.

65. A kit for detecting a SLC5A8-associated cancer in a subject, comprising at least two primers of claim 64.

66. The kit of claim 65, further comprising a compound to convert a template DNA.

67. The kit of claim 66, wherein the compound is bisulfite.

68. The kit of claim 67, wherein each primer comprises at least a CpG dinucleotide.

69. A method of converting a nucleic acid sequence at least 95% identical to any one of SEQ ID NOs: 12-13 or fragments thereof, to a bisulfite converted sequence comprising:

a) providing a nucleotide acid having a nucleotide sequence as set forth in any one of SEQ ID NOs: 12-13 or fragments thereof; and

b) adding a bisulfite compound,

whereby the unmethylated cytosine bases of the CpG islands are converted to a different base.

70. The method of claim 69, wherein the unmethylated cytosine is converted to a uracil.

71. A nucleic acid sequence as prepared by the method of claim 69.

72. An isolated or recombinant methylated SLC5A8 nucleic acid, comprising a nucleotide sequence as set forth in any one of SEQ ID NOs: 12-13 or fragments thereof, wherein the cytosine of the CpG island is methylated.

73. An isolated or recombinant SLC5A8 nucleic acid, selected from the group consisting of:

a) a nucleotide sequence as set forth in any one of SEQ ID NOs: 12-13 or a fragment thereof;

b) a complement of any one of SEQ ID NOs: 12-13;

c) a nucleotide sequence that hybridizes under stringent conditions to a nucleotide sequence of any one of SEQ ID NOs: 12-13;

d) a nucleotide sequence that is at least 98% identical to the nucleotide sequence of any one of SEQ ID NOs: 12-13; and

e) a nucleotide sequence comprising at least 50 consecutive base pairs of any one of SEQ ID NOs: 12-13,

wherein the SLC5A8 nucleotide sequence is differentially methylated in a SLC5A8-associated disease cell.

74. A method for detecting colon cancer, comprising:

a) obtaining a sample from a patient; and

b) assaying said sample for the presence of methylation of nucleotide sequences within at least two genes selected from the group consisting of: SLC5A8, HLTF, p16, and hMLH1;

wherein methylation of nucleotide sequences within the two genes is indicative of colon cancer.

75. The method of claim 74, wherein the sample is a bodily fluid selected from the group consisting of blood, serum, plasma, a blood-derived fraction, stool, urine, and a colonic effluent.

76. The method of claim 74, wherein the bodily fluid is obtained from a subject suspected of having or is known to have colon cancer.

77. A kit for detecting colon cancer in a subject, comprising primers for detecting methylation of nucleotide sequence within at least two genes selected from the group consisting of: SLC5A8, HLTF, p16, and hMLH1,

wherein the primers for detecting methylation of SLC5A8 nucleotide sequence are selected from SEQ ID NOs: 5-11;

wherein the primers for detecting methylation of HLTF nucleotide sequence are selected from 5'-TGGGGTTTCGTGGTTTTTTCGCGC-3', 5'-CCGCGAATCCAATCAAACGTCGACG-3', 5'-ATTTTTGGGGTTTTGTGGTTTTTTTGTGT-3', 5'-ATCACCACAAATCCAATCAAACATCAACA-3', 5'-GCACGACTAAAAAATAAATCGCCGCG-3', 5'-AAACACACAACTAAAAAATAAATCACCACA-3', 5'-TAAAACCTCGTAACTTTCCCGCGCG-3', 5'-GTCGCGAGTTTAGTTAGACGTCGAC-3', 5'-TCCTAAAACCTCATAACTTTCCCACACA-3', 5'-AGTTGTTGTGAGTTTAGTTAGATGTTGAT-3'

wherein the primers for detecting methylation of hMLH1 nucleotide sequence are selected from 5'AACGAATTAATAGGAAGAGCGGATAGCG-3', 5'-CGTCCCTCCCTAAAACGACTACTACCC-3', 5'-CGTTTTTTTTTGAAGCGGTTATTGTTTGT-3', and 5'-AACGAACCAATAAAAAAAAACAAACAACG-3'

78. The kit of claim 77, further comprising a compound to convert a template DNA.

79. The kit of claim 78, wherein the compound is bisulfite.

NCBI GenBank Accession Number AC063951
Human Genomic Sequence that covers 15 exons of the Hui1 Gene and the promoter region. Base Pairs 82200-83267 are highlighted in yellow on page 35.

GAATTCTTTCAGGGTCCTCCACTAATGCAGGATAAAACCCAACCTAACTTTTC
AAAGCCAAGCCAGAGCTTTGCTGATATCATGTCATATCATGTGTCCATGCGCT
CAGACTGCAAACATTTATGAACACACGTCTTTTCTGTTCCCTCTCTTAGTATTT
TTAACCCCGCTTCTGACTGGCAGACTGCTTTTCAACCTTCAAGGACTATCCCA
AATACAAGTCCTTTGAAAAGCTTTCTCTGTTCCTCAACCCCAGGCAGAATTAA
TAGTTATAATTCAATATATCTCCAACAGACAAATTCGCTAGACTATAACTACT
CGGTTCCAGGCTCCATTTCTCCAACAACTATCAAAACCTGCAGGCATCACTGT
TTACCCTGGAATAGCCAGCTCATACCCAGCACACAGTAAGCTCTACTGCTTAC
TGCAATATGTGCTTACATATATTGCAGAATATGTATAGATTTAGCTTTTGCCA
CACTTTATTTTGCTTATTTGCTATATAAATTAAGCTTTTATCACACTTTATTTT
GTTTATTTACATTACTGCATGCAATTCCCATTTCCTTAACAAGATTATAAATTC
CTTGAGGGCAGAGCCTTTGTCACACTTCACTTGTATTCCTCAGAAGGTTACTC
AATGCTAAACATATTATTAACTGCTCAATTAATATTTGTTGAAGTAAATTACT
TTTTCTATGGTTTCTTTGATGAGGGTCTCCGGCAAGGAAACATTTTCACCCAG
GATCAAAGCAGAAATTACATCCAAGAGGGTCTCCCAGCAAGATGTGGAGAG
ACTGGCTACTTGCAGTGTTTGAGATAACACCTGCAAAAAGAAAAAACAATGC
ATACTCAGCAATCAGCTTTTAAAAACCAGACTCTCGAGCATATTAAGTTGGG
AACTCACCTTACAGACATCAGCAGGCGTGGGTATCTTTGTCCCACTTCCATGT
TTTACAAGTATAAGGTATGTTTCCAACAACCTTTTAATCTGTTCAGAACTTTC
ACAACAGTTAGTTTTTGTTACTTTTGTGTGTAGATCCAAGAGCGATTCCTGCA
AACAAAGAATTTAACAGAGAGAATATAAAGTATCAAACTCAACTGTTAATCC
AACAATAAAAACTGTGGGACAGTTCAACTTTTCCCATTCTCTTGCTTTGGAAG
AAAAAGATACATTTACCAGGACTCACTAACGATAACAGCTAACATTGGCTGA
GTGCATTGTTCTAAGTCTTTTTTTTTTTTTTTTTTTTTGAGATGGAGTTTCACT
CTTATTGCCCAGGCTGGAGTGCAATGGCGCAATCTTTGCTCACCGCAACCTCT
GCCTCCCAGGTTCAAGCGATTCTCCTGCCTCACCCTCCCAAGTAGCTGGGATT
GCAGGCATGCGCCATTTTGTATTTTTAGTAGAGACAGCGTTTCTCCATGTTGT
TCAGGCTGGTCTCGAACTCCCAACCTCAGGTGATCCACCCACCTCGGCCTCCC
AAAGTGCTGGGATTACAGGTGTGAGCCACTGTGCCCGGCCCATTCTAAGTTTT
TTACAAGTATTCACTCATCATCCTCACAGCAACCCTGAGGGAGAGAATATTA
CTACCCCCATTTATTATCTGAAGAGACTGGGGAATCGAGATTTCAAATAATTT
TCCCAGGTTACACTAGCAGTAAGTGGAAGGGTCAAGATTCAAACAGGCAGTC
TGGCTCCAGAGCCTTCTACTGCATCTCAAGATAACATATCAAATAAAAAATA
GCACAGGGGGGCAGAGGGAAGGAAAATTTTAATATGTGTACAGAAGTATAA
ATAAAACAATTATAAAATATAACTTCTAGGAGTGATTTGCAAAACTGAGAGC
AGAAACAGTAAATTCTTAGACTCCTTATCAAAAACCCATCCTTAAAGATTAG
AACTCACTTGCAAACATTCAAAAAATGTACCAAAATGTTCCTTGGAGATGTA
GGATACAGTGGATTTGACCATGTTTTTGAGTGTTTCTCCAATTAACATCCATG
GTAGTTGAGTTTCTGTTTCAGTGACTGGTCCTAGCTTTCGCAAAATTAGCTTC
ACTGCCTGTGGGAATACAAGTCATTAACATTAATATATAAACTCTTATTTTCA
AGTATTGCTCTGGGAGTCACCACTCTGATTACATTAGCTACCTTAAAAATCAG
GTCTGGAATAGTCAAGGATATGGCAACTTTGATGGTTAATAACGTGTCCCAA

AAGGAACTTCTAAACTAGGATGGGAATTTTATCATTACAACTGTTCCCTCCCC
ACCTGCCCTTGCCCATTAATCAATTTATCTTACCCCCTAGCATCCAAAATAAA
GAAGACATCAATGGCTCTGTAAGTAACATCTCACTGTCCAAGGCCTACATTG
GGTAAGCATTGTTACTGAACACCCAACACTCCTCATCTAGTGCTATGTACTCT
AGGCTGTCTTCCCTCCACATCATTTGCCTCATTCTTTTCTACTTTCTGTTTCATC
CATCCACATAGACTCTAAGCACCCCTCCCCATTATGTACCTGGCCTGTACAGG
AGTGAAACATATTTCTAACTCCTTTGCACATTTCAAAGAGCAACTGTCCAACA
CCTTCAACTTTTTCTGGATGTTTATCAAGATCAAGAAACATTAAATTGAAAAG
TGCGTTTTTATCAGAGACCTAAAATATAAAGTAGAAATGGAAAATTATATTC
ACAAACATAAACATATCCAAATGAAGAACTCTCATTTGCTATTCTATATATTG
TTATTTAAATATACTTAGGAAAAAATTGTGAAAAGAACAGACATTTGTAATG
ATTTCAAACTACTCTAATAGTGTGTTCTCACTTGGAATGTCACACAAAAGAGT
GAATTCATTGGTATTACAGGAAGATATGTACTACAAAAAAAACTACAGAAGT
CAACATTCTATCTTTTAAAGTAAATACATTTGTATTTTGTATTGTATACAGTAA
AATATTTGTATTTTATGAAATAAAAAATTTAATATTAATGTTTTTGTTTTATTA
CAAAAACCCTCTAAAATTTACTATGAATATTTTATACAAACACAAATGACAC
CATGCTACCAGTTACTTCCAGGAAACTATTAGAAAAGTTCTTGTTAATGTCAC
AGTTCATGTATTTACTGACTTGAATTAGAGAGCATTTTTGATATATACACACT
GAAAGCTATCACAAACCTAAATAACAATACAATAGTGACCCAAATTAGATCA
CATCAATTTGTGAGCAACAAATAAGTCTTACTGTGAGATAAATTCAAAATAG
TAACAGATTTCTGTTAAGTATATATATAGATACATAGAAAAAATTAGCCCAT
AATAACCACATTTTTAAAAGATATCTACTTCTATTCATCAGATTATTATTATTA
TTATTATTACTAGTGTGTGTGTGTATGTGTGTGTGTGTGTGATATGGAGTCTTG
CTCTGTCGCCTAGGCTGGAGTGCAGTGGTGTGATCTCAGCTCACTGCAACTTC
CACCTCCTGGGCTCAAGCAATTCTCATGTCTCAGCCTCCCAAGTAGCTGAGGA
TACAGGTGCGCGCCACCACGCCCAGCTAATTTTTGCATTTTTAGTAGAGACGG
GGTTTTGCCATGTTGGTCAGGCTGGTCTCCAATTCCTGACCTCAGGTGATCCA
CCTGCCTCGGCCTCCCAAGGTGTTGGGATTACAGGCATGAGCCACCACGCCT
GGCCTATTCATCAGATTCTTGATGATTAGCAACAAACAGATAAAATACCAGA
CTAACCTTTCTCATCAAAAAAGTAAAACTTTCAGCAGCAAAATTTCTTATATG
TAGTTTTTTATGAGCCAGGAGTGTGCTGTACATGCTATACATGAAAAAAATA
AGATACATTTCATTAATCATATAATTGTAATAAATACATACTACATGTCAACA
ATATGGGCAACAATGTGCTGGGTATGCAAGGAATACACAGCAGGTATCAAAC
AAATTTAAAATCTCATTCATTTATGGAGACACCCACATGTTGAAAGGAAGAC
TTGACCACAGACATGAAGAGTCCTAGGACTGGTGGTACTGGTTTTACAAACA
AGACTCCAGGAAAAGTTGAAATTTGTAATGAGCTCTGAATGAAAGAAGAATT
AGGTGGGGACTGCAGTCCATATTATTGGTATAAAAGCAAGAGCAAAGATGAG
GCAGGTGGAAATGATCATGGTCATGACAAGGAGGCTGGTCCATCTAAAAGAG
GAAAGATGATACAGTAGAGGAGAGCAGCTATGGATAAAGTTGGTCAGGTAG
ACAGGTCTAGCTTACATCTGTATAAGCACTTACTCTGTGTTACGCCATTTAAT
CAGCACAATAACTCTATGGGATGGGTACTATTATAATCCTCCCATTCTACAGA
TAATGAAAGTGAGGCAGAAAGCATAAGCAACTTGCTCAAGGTCAAGCAGCC
ATGCATCTATAACTAAATAATTACTTATATATAATCACATTGTTAAATTTGGT
CTCCCTAATGATAGAAGGGTATGGAATATATCTCTCCAATTTTCTCATAACCC
CAGTACCTAATAGTTCCTTGCTGACAGCAGGTACTAATAAATGTTGGCTGAAT
GAGAAATGACCATTTTCAGAAAGACTAATTTGGCAGCAATATACAGGATAAA

ATAAAGGAGGAAAGAAGAGTCTGCTAATTCAGTCAGAAAGGTGCTCAAGTC
ATACAAGCTTGGGCTAACAGGCATGAAAGAGACTGGAAGGAGAGGCAAAAT
GGCAAGGGATGAACCCAGTAGACATTTCAGGAGTGCCCACAATGAAGCTGA
AGACCTTACAGCGGTCCACAGGGCCCTGGATGATCTGGCGCCTTACTACTGCT
TCTCTGACATCACTTAGCTAGTCTCACCCTTATACAGGCTGCTCGGCCACCTA
AAACTTCTCCCGGGCATGTGCCCAAGACATTCTCCCCCTGCTAAAATGTAAAC
TTTGTGAGAAGAGCATTTTTTATCTGTTTTATCTGTGACTGTATCTCAAGTGCC
TGACATACAGACAGTGTCCAATAAATATTGACTGAGCAAATGCATGAATGAC
AGAATCAAGAGGATGTGGACACCTCTGAAGACAACTGGGTTAGTGATGCCTC
TGATATTCCAAAACTCAGAGACAGGAAGAACGTTAGTATTAATGACAGAAAA
TGGGCAACCAAACAGGGCCTGGCATACGGCAGACACTCAATACCTATTTACT
GAACACTTGAATGAATGTACAGATAAGAGGAGTTGATTTAACAGGAAAGTGT
TGAGTTCAGTTTCAGTAATACAAGTGAAAATAGCCAGGCTACTGAAGTTGTC
ACACCGAAAGCAAATAAATTTAGAACTGAAGAGCCTGACTTAGGAGACATA
AAAGTGGCAGCTTTTATGACAGCTGAAGTCATAAAACTAAATCAATTATTCC
CCAAAGACCTTGTTGGTTGTTTAATGAAATAATATTGGTTTCTAACATATTTG
ACCAAAAGCTTCATAGGGAACTAGTAAAGGCTGGAAAAAGATTTCTTTCTCT
TTCTTCATTCAAAGTTCCTAAAAGAGAACGGTGGGCTGAGTGAAACGGGTGG
CATAAAACAAAGTCTATTGTCCCATACCATGCATACACCATTTACTGTGGGGT
AAGGGGCTGATGACTGATATTAAACACTCCTGTGGCAATTATAGAACATGAA
AACAACCGGAAGACCAAGTGGAAAATCCCACCAAACCCCAGACTTAAACAG
ATAATAAGACAATTTATGCTTACTTTAAACAAAAGGTTACTACAAAATTCATT
CTCTTTCCTTGTCCACATTTACTCACAAATACTGAGTGTCTCTTTCACCTTAGT
AAGTATTAGTGGCTGGGGAAATGTCAGTGTTTAAAACAGGGCTCTTAACCTG
TCTTAAACCTCCCTTATACCAGGTGTAGAGAGGACAGCAATACATTATAACC
AAACAATTCTGGTAAAAGTTGTGTGACATAGGTGGTTAAAAAATGCCAAGAG
GATAGAGAGTACAAAATTTTATTAATACTTCTGGCAAAGGTGCCTGTTTGTGG
AGGCTGTAGGGACCGAGTCAGGCCTTCTTGGGTGGGTATACTTCCAGTAGCT
GGAGAGGAGAGATGGTCATGGCAACTCAGACTGGCAGACTAGATTGTGGAC
CACTTTGGCTGGAGAATCAACTTTGTACATGGAAACAATTAGGCACTGAGCT
GGCAAAGTGGTTTGGAGTCATAAAGAAGCCAGGGAAAAAGAAGCAGGCAAT
TAAAAAACATACTAGAGAGAGAAAGGGGTTACCTGTATATACTGGACATGTC
CTTCACCATCAGTCTCCACAGGTACTTATAAAGATATGATAACGAGGTGAAA
GCCCATTCTAACAACTCTGTGTCCTGAGTCTCCAGGATCGAGGTGATAGTCAA
AAAAAACTCTGGAAAGTGTGGGTAGAAATCCATCTGCAGATCTCGTGCCAAC
TGTACAACCAAACTGGATTAAAAAAAGAAGCAACTGATCATGTGGATTTTTT
TAAAAGGCTACAAATCTACTTAACGATAGTAGGTGGCCTATGATCATAGGAT
TAATTAAATAAAAATACTTCTAAATCAAGTCTGAAAAATTATAAAATTCTTTA
AATCTTAAGATCTCAGCAAAGAAATCAGGCAAGAAGTTACATTCCATCATAA
GTTGTAGCCAACTGTTCCTATAGTGTCAAAGAAAAATGAGATTATCTTAGTAT
ATAACAAATGCGAAATTATACCCCAAGATTCCACTTTATGGTAATGCACTTAT
TCCTAACAGAAAAGGAAAATCCTCCTTGTTTCTAAATTATGAGCACCTGATCT
GCGGTATATGTAGCTTTGAAACACAATCTTTCTTAAAATTCAGGCAAATATAG
TTGTGAGCTTTCTCTCAAGATGCTACTTACTCCAAAAGGGGTTGATAGGCAAA
ACTGTTCTTAACTTGCAGGTGAGTCTTCAAACTCTGAACTATCTCGTTTTGGT
GATACACCAACTGATTGAATGATTGGCATTTGTCAATAACTTCTTTGTAAAAT

TTTCCTGAGTGGGGGAAAAAAAAATATGTTCTTTTCTATCACACCACTATTCA
GCTGTATGGAACATATAAAGTGCAAACTATGGGTGAATAAAATAGAAACGA
AATAGAAAATACCGAAGTGTTCTGTGAGGTTTAATTCTCTCCATTTCAGCAGA
CCCTCAAAAAAGTAGGTTTCAACCTCCTGTCAAGATTAAAGCAGTCCATTAAT
CAAATAACTCTAAGAAAGCAATGATCAAGGTAAATAGAATAATGATATATTA
ATATCTTAAGACAATGATACTGAAGAGTTTAAGAACTCCTTTTAGTTTTTTAA
AACAGGCCATTAAGTAACACACTAAAAAGCAATCAGAAGTTATCAATAGGCC
ACTTATAAAATGCTGTTTTGTTGATTTGGTTTCAGCAAATAATTTTCTGGGGC
CTATATATATTGTCCACTAAGGGAAAATTATTTTCTGTGTTTTTTTATAATTTT
AGGCAAAGAATAAGGTGTTGGTTCTTTACAGTTCCTTCATTCTGCTTTTAGAA
ATATGAATTACTACAACCTTATAAAGAAGTAATATGGCATTCCTGTTAAAATT
CAAAATAGTTACGCTCTTTGACCCAGGTAGTTCTTATAAAGTTTGCAAGCCTT
TAAGTAAAAGATGTTTATTCAGCTTAACTACAGTGTGGGGAAACATTAAACA
GGCTAAAATATCCGACAATAGGAAAATGGTTGGAATAGCCCATGGTCTATAT
ATACTAAGGTATATTATGTGGCTACTAAAAAGAGCTATATCTCTATTGAACTA
GAACTACACTGAAGATATGCCCCCAAAATGATAAAAGTTATGCAGTGAGAGG
GTATGGCAGGATTCAATTTTCCTTAAAAAACAAAAATAAAAAACCCTCTATA
AATGTTTGTACATACGAACATGAGAGGACAGTATGGAAGAATACAACTATAC
TGAACTGTCAGTGTTGGCTTCCTTGGGAATTAGGGGTGGGAGGAGTGAGATA
ATGGGCTTTTCGTAAGTTTACCACTATGTTACTTAACTTGTTAACATGGTATC
AATTACTTTGTACTTTGAAAGGTAAAGCCAATAAATTATCATACATGTATGAC
ATGTATGTATACATGTACGTATGTGTGTATATAGATGTATAAATAAATGCATA
CCCCCCCAAAACAATCATTCTACCAAAAAGATACATGCCTTCGTATGTTTATC
TCTACACTATCCACAATAGCAAAGACATTGAATTGTCCCAGGTGTCCATCAAC
AGTAGATTGGATAAAGAAAATATAGTACATATACACCATGGAATATTATATG
CCTTAAAAAAAGAATGAAATCGTGTCCTTTGCAGCAACATGGATACAGCTGGA
GTCCATAATACTAAGCAAATTAACATAGGAACAGAAAACCAAAAACTGCATG
TTCTGACTTATAAGTGGGAGCTAAATATTGAGTACACATAGACATAAATATA
AGAACAATGAACACTGTGGATTACTAGAGGGTGGAGGAAGTGGGGATGGGT
TAAAAAAAACTACCTGTGGGGTACTATGCTCACTACCTGGGTGACAGGATCC
ATACTCCAAACCTCAGCATCACACAATATTCCCATGTAACAAACCTGCACAG-
GTACCCCCTGTATCTAAAATCAAAGTTGAAATAAAAATAAAAATAAATACAA
ATATGTGTTTATAGAGAGAGAGAAAAAAGAGAGAATAAACACATAAGCACA
CATGCAAACAGCATGCCAAATCTACAATATCAAAAAAAAAAAATCCTTAAACT
GTTCTTTGGAAATCTTTAAAATCAATAGTTAGGCAGAATAGATACTATGTAAC
CACAAATATTAAAAACTAAAAATTAAAAAAAAAGGCAGAAAAGAAAGAGAA
TCCCATTAAATTTTGTTTTAGGCTGGGGACAATGGCTCATGCCTGTAATCCCA
ACAGTGGGAGGCTGAGGCAGGAAGAGCATTTGAGCCCAGGAGTTTGAGACT
AGCCTAGGCAACAACACGAGATCCTATCTCTATTTTTCAAAAGTAAAAATATT
AAAATTTTTTTGTTGTTGTTTTCATGTCCTTTAAGGCATTTTCATGTCCTTTAA
GGCAGAAAGAAAATATGCAACACAGTTTAAAACTTAAATGCAGAACGCATTT
CTAGCCTAGCACAGACCTGGCGTATGTCAGCTATGTGTGAGAGACCATGTCA
CGTCCTTTTGCAAGGTAACTCTGGAGCTCCTTTCACCAAGAGACGGAGTCTGT
TTCCTCACCCTTTGAATCTGGCCTGGCCTCCTGACTTGCTTTGACTAATAACAT
GCGATGAAAGTGACTCTCATGACCAAAGCAAAACTTTATGAGGTCTTGACAG
CTTCTGCCTTCACTCTATTGAAAAGATTCTGCCACCATGAAAAGAAGCCTAGT

CTAGCTTATTGGGGAATAAGAGGCCATGAAAAGAACCAAGATCAACAAGCC
ACAGCCAACTGCCAGACACGTGACAGAGGCCATTCTGGACCATCCAGTCCAG
GATTCTAATTTCAGGCTGTCTTCTCTGCCTTCTTAACACTGCCTCTTGATACCT
TCAACCTCTCCTGTGGCTTAAATTATTAATCATATTCTAAATGCCAAATCTGT
AATTTCACCTCTGAAGTCACCTATACTCCAGATCCATATATACAAACATCTTT
TGGACACTATCATTTGGATGGCCAAAGGTATTTCAAATTCAAGCCCCAAATG
GAACTAATTATCTTATCTTTACAACCTTGTGCTCCTTTACCTATAAATATATCT
TGATGAGTAACATCCTCAACCATTCAGCCTTAACCTCATTATCTTCCTTGCTG
CAAACCCAGTAGGTCACAAGTCCTATCCATTCATTCTACTTCCTCAACATCTC
TGGAATCTTTCTCTTCTCTCTATCTTGATTGCTGCTACCATAATCATTTCTTAC
TTAGATAACAGTCAAATCCTCACAACTGGACAACTGAGATTGGATCTCCAAG
TTAAAGGCATGTTACACAGGACCTTAAAAGGCTAGCAGAGGAATCATGACAT
GATGAAGAACTGCAGTCCTGCTCTTTTCTCCACACTAAAGCCAGAGAGATTG
AAAATTTAAACCTGATCATGTCATTCATTCCCCTGCTTACAGTCCTAGAACTC
TAAGTCCCTACAGCTTCAGGATAAAACCCCAACTTAGCTTCACATATGAAGTC
CTTTATGATCCTTGCCTATTTCTCCAGCTTTATCTCACCAGTTGCCCCCTTGCC
CCATCCCAGTCACCATGAATACAAAAAAAAAAACACCCTATACACACCTATAA
CCACACTGGACTGCTTTGCAGTTTCTTGAATGTGACATGCATTCTCTAGCTTCT
GTTCCTTTATATAGGCTTCTCCCTCAGCCTGGAACAGGCTTTCCTGTCCTCTTT
ACCATGCTAAAACCTATATAAACAAAATTCAGGTGTCACCTCGTCTAGATCA
CCGTTTTTTTGACACCCCTAAAGAATGTAGGTTAGGTTCCCTTTACTATGAGT
TCTCAAAGCACAATGTGCTCACCTAGACGACATCATTTAACACATTGAATTTA
CACAGTGGAGATAATGCATTCAAAGACCCAGCGAGTGGAGTCTGATTATAGT
AAAGAAGAATATTACGAACTGAGATTGGATCGCCAAGGTATAGGCATGTTAC
GTACGACCTTAAAATGCTAGGAGGAATAACGACATGATGCAGAAAAAGACTT
CTGCATAAAAGGATGACAAGGCCGGGCGTGGTGGCTCACACCTGTAATCCCA
GCACTTTCGGAGGCCAAGGCGAGCAGATCACGAGGTCAGGAGATTGAGACC
ATCCTGGCCAACATGTCAAAACCCCGTCTCTACTAAAAATGCAAAAATTAGC
TGGGTGTGGTGGCACGTACCTGTAATCCCAGCTACTCGGGAGGCTGAGGCAG
GAGAATCGCTTAAACCTGGGAGGTGGAGATTGCAGTGAGCCAAGATCACGCC
ACTGCACTCCAACCTGGAGACGGAGCGAGACTCCATCTCAAAAAAATGAATA
AATAAATAAAAGTAAAATAAAGGGATGACAGAGTCAGAGTATGGTTAAAAC
TGGAAAATATTATGTAGTCTAGCCCCTTTATTTTTATAAAGGAGAAAATTAAG
CCCCGGGAAAAGGGCTTCCTCAAAATCACTTTAAAGTTATAGCTTCAGGAAT
ATGGATCTGCAGCAGTGCTTGGAATGCATAAGGGAAAGGGAGAGGCTAGAA
TCACAAAGACAGCTGAAAGTCAAGTCAATTGTCTAATAGAGCTTCACCCAAC
AGAACTTTCTGCAAAGATGAAAATGTTCCAATTCTATATTTATTCAATATGTT
AGCAACTAGCCACATTTGGGCACCCAAATTTTTAATTTACTTTAAATCCAATT
TGTCATATGTGGCTACTGTATGAAACAGCACAGGTCTAAAGCATTTCATGTCC
AAAAAGGAATACCTTGAAAAACAATTCACTTCTACTAACAGAAGAAACTAAA
ACACCATGAACACTTGAAGATTGACTAGTATCACATTCTCTTACCTCCTCATA
GCTTGCAGTTCTATCAATCCGGTGAATAATATCAATATTAACATTCCCCAGTC
GTTCAGCAAATGTAAGAAACTGGAAGGGAAAGTATATTTAAGATACATAATT
AATTAAAATTTATCAGATCTTTAATATCTATTTGAATGCTGCATGTAGGCATC
TCTAATCACAAAGGATAAGTGGAAAAATAAACTGAAAAACATACGGCCGTA
AACAAATTTACTGCATCACTGTTCAAAGATAATGAATACTTCTATGTTTGCAT

AATTTCTCTCAGCTATGTCATTTCAAATAAAATTTCCATTGCCAGACTGGTGA
GCCTAGGTGGATGCTGGCAATTAGTCTCGCTAGATCTATTAGGTTTCATACCC
TCCCATAAGCATGGGGACCTAGCAAAGTCGCTGCAATAAAAGTGTTTTTAAA
CATATACAGACCTATGATTGTATCCTAAGGAAGACCTGGAAACAATCTATCA
AGGGGCAAACAGAGAAAGCGCTGTATATTTGCCCTTAGCTGGGAATCACTCA
CCGCCAGCCGACTCGCCCAATTCGGTCTTTAAAGATAAAAGAGCAGGGGAGA
ATTGGTCCTAAGCAATCTCCTGGAATAGTGAATTTAATTCTGGACTACAGGAA
ATTCCCAGGACTGGCCAGACCCCATAAAACATGGGTGAAACTTGCTGTCCAC
ACTTCTTTCCTCCTCCAACCCATGTTTACTACATCCAGTGTCTCCCTCTTTCGC
CGGAGCCTCCAGGAAAGTGACACACTCGGCCCAGAAGTCTGAGGCCCCTGGA
GTCTCGCTCAGAGCCTGTCTCACGACTGAGGCAGCGGAGACCCGCGGCTCCC
TGCCTAAGCTCCCGCGCTCACCCGGTAGGTGTTCTCGGTCTTGTGGGAAACGG
GCTTTGTCTTCATGGCTGCAGAGGGCCAGTGGCCCGCGACGGCCTCGGGAGT
GTCGAAGGGATGCAACCGACAGTAAGGAGGGGAAAGCGGCTCACAGGCTAT
ACTCTCCGATTCCCAGATGCCCAGACTTTCTCACGTGCGGCTTGAGCCCCTGG
GCGCCGCCATGTTGGAGACAAGGAGGAGCCTGAGTGGGTCACGTGGACGGA
AAAAAGAACGGCGCAGGCGCACCCTTTGGTGGGGTGGGGCCTACAGGAGGC
GGGGCTGCGCACATAAGGGCGGGCGTTTGGGTGAGGTGTTCTTTTCACTCCCT
TCGGTAAAGGTTTAGAAGACAAATGTATTTTCATTATAAAATAAAACATACC
TGTAATCGTTATCAACTAACATTACTGTCCCTCACTACGTACCTGCATCGTGC
AAAGATCCTTTCATCCATAATTTCACAGTAAAGCTTATTAGGGATGTTAATAC
AAAGGAGGTACTGCGTCTATCTATATATCTATATATAGATATATACTTTTTTTT
TTTTTTTTTTGAGACGGAGTCTCACTCTGTCTCCCAGGCTGGAGGGCAGTGGC
GCGATCTCGGCTCACTACAATCTCCGCCTCCCGGGTTCAAGCAGTTCTCCTGC
CTCAGCGTCCAAAGTAGCTGGGACTACAGGCACCCGCCGCCACGCCCGGCTA
ATTCTTTTGTATTTTAGTAGAGACGGGGTTTCACCTTGTTGCCCAGGCTGGTC
GCGAACTCCTGAGCTCAGGCAATCCGCCCGCCTCGGTTTTCCAAAGTGCTGG
GATTACAGGTGTGAGCCACCGCGCCTGGCCGGTACTGTGTATATTTTTAAGCC
AATTTGACACAAGAGGAAACCAAGGTTGTGCAGCTGGTAACTGGCAGTCTTC
ACTCAGACTCTAAATTTTGGTATTCTTAACCACTACGCTGTAAATAAAATTGA
AAAATAGAGAAGGAGTTTTAAGAAAAATCTTTTAATTTCATGCACAACCGCT
GTTAATGTTTGGTGTTTTTCCCATCACTTTAATGTCTACACATAGCATGTTGTT
AAATTGTACAGTTATATATACATATATGTACAATTTTAAAAAATAATTTAAAA
ATATAATTATTTACCCCATTTTACATTCATTCATTCATTCAGCAACTATTTCTT
GAGAGTCCTCTGTGGACCAGGTACTGTTCTGGGAGCCAAACGAGAAAGGCAA
TAATAGTATTAATTAATTATTTGAAAATATTCACAGCCTTCAATTCCTGGGCT
AAAGTGAGCCTCCCACCTCAGCCTCCCAAGTAGCTAGGGCTACAGGCATGCA
CCACCACACCCAGCTCTTTTATTATTATTATTATTATTATTATTATTATTA
TTATTTGTAGAGATGGGATCTTGCTATATTGCCCAGGCTGGTATCAAACTTCT
GGACTCAAGCAGTCCTCTTGGCTTGGCCTCCCAAAGTGCTGGGATTACAGGG
GTGAGCCACCATGCCCAGACTGGAAATACTTGGTCTGAGATGCTTACATTTTA
GTCAAGGAGTGATGTTAAGTAGACAATTGGATATATTGTTGTACAGAGCTCA
CTGTAGCAGTCTGGGCTGCAGACGTGACTTAAATACACCTAAATACCTCAAG
TGTGAGGCATTTGCGACTGGGAAAGAGAGCCAGTAAGGTAGAAACAGACTA
GTTATGGCATGGAAGCAGAGCTGGTGGCCAACTATACTGCAAATTCGAGGAT
AGAGAAGTGGTTGCTGGATCTGACAAGGTGAAGGTCACTGGTGACCTTGCTA

AGTACAGTCTAGTTGGAAGAAAGGAAGTGAAATATGCATGGATTGAGGAGG
TAATCAGAGTTGAGAAAGTGAACATAACACTTTCAAATTGTTATTGGCATTAT
CACATCAACATTTTCTGTGTCACCACAGGGTTGTTCTAAATATTTTAATGCCT
CCATAGCATTTATAGATTTTTTTATAATAAAATATTGACGTGCCTAATGTTCC
ATTGGCTTTCAGTAGGCTAATAATTGGTCTTTTTTTTTTAAAAGCCTTTTATGG
GTCACAGGGAAATCTAAGAAAGCTGTAGAACTTCTCCCAGAAAGATGCAAAG
AAGCTCATCCACACTAAAATGTGCACGTTTCAGAGAGTTAGAAGCCTGCAGT
TAACTGAGGGTAGAAGCCCCCACTCAAAGCAATATTTTCCCAACCTGTTGCC
AAAATCACCCCCACATCATTGTGCCATTTTCCTTCTCTGCAATTTATTTTTGGT
AATGGAGTTAACTGGACAATAAGGAGTGAAGAGAAAAATCAAAGAAACTGA
AAGAACTGACACAAGTTATATTGACCATTTATTGTTTGCATTGTTCTTCATTA
CATGACCCTCACACACTGACTTGACTCTGACTGGCTTAACAATATGTCAATAG
AACATGGTGGTTGAGAGCATAGATGCTGGAACCAGGTTGTCTGGAGGTAATC
CTGGTTCTGCCATTTATTAGTAGGGTCAGTTACCCTTTCAGAATGACATTTCC
GTATATGTAAATGAAGCTAACAAAAGTAGCTACCTCACGGGTGGTGATTATG
CAGAGTAAATAAGCTAATGGATAAAAGATGCTTAGAATGTGCCAGACACCAT
AATTGCGCAAGCTCAGAAATTACTTAACCTTTCTGAGACCCAATTTCCTCATC
TATAAAATAGAGATGACAATAATACCAGTCTTTCAAGATACTGTGTTGGTGC
ATAGAGCATGAAATAAATGTTTGCTAAGTGGATAAATAAATGACCATTTATT
CTCTACTACCTGGCTACCTGACAAGTTCCTTAAGGGCTGACACCTTATACTTT
TTCAACTTTGTGTCATCAGTTCCCTTGAACATAAGAGTTGTTTAATGAATGTC
TGTTGAATTAACAAAGCTTATAATACTTACATTGACCCGGACAAATTTAAGA
GTAATATAGTTCCAGCATTGGATGTGAATTGCACCACATGATTTTTGGTATCT
CTCAAAACTAAGTTTCTAAGCTTTAATGGAATACCGAAAGCATCCGTGAAGA
GCACGAGTGTTTCTTATAAACATTTCCTTGCTTCCAGCGTTTGTCCTCTGTCAA
GGTATAGGATGAATATCCTATACCATCCTTTATATCCTTTTTCCATTAGGGCTC
TGCCTGTCCTAACTACTCCAGCCAGTAAATAACATACATTATTTTTTCATTCTT
TTATTCTTAAGCTTTATAGGGCGCTCCTGGAAGTTTTGCTTTTTAATTTTTGTT
TCCTTATTGCTTAGCGTGCTGCAATTTCAAGCCAAGAAACTTTAAGAGCATGT
AAGAGCTAGGCCCAGTGGCTCATGCCTGTAATCCCAGAGCTTTGGGAGGTCA
AGGTGAGAGGATTGCTTGAACCCAGAAATTTGAGACCAGTCTTGGGCAACAA
GGCGAAATCCCATCTCTACAAAAAATACAAAAATTAGCCAGACATGGGGATG
TGCACCTGTAGTCCCAGCTACTCAGGAGGCTGAGGTGGGAGGAACACTTGAT
CCAGGAGGTCAAGGCTCCAGTGCTGTGATCCTGCGACTGCACTCCAGCTTGG
GTAACAGAGTGAGACCCTGTCTTGAAAGAAAAAAAAAGGAAGGAAGGAGAGG
GAGAAAGAAGGGAAAGTAAAGGAAAGGAAGGGAGGGGAGAGGAGGGAAAG
GAAGGGAAGAAAGCACATGAGGAGTTTACCCAGCCTAGACAAGAAAGTGGA
ATCCAGAGAAGGCTTCTTGGGGGAAGTGACATCTAAGCTGAGACCTGGAAAA
TGAATAGGAATTAGCCAGGCAAGGAATGGACATGAATGGTGTTATCTAGGTA
GAGGGAGGAGTATAGGCATTTGTCTCAAAATGGTTGAGACAAATTTGGTTAC
TTTTAGTTTCCAAGGCAAAGCCACACCCTGTCAAATTAGCATTGGCCATGGGT
ATCATCTTTAGCATCCTTGGACTAATAATAGGAAAGAATAGGGACAAGATGA
AGCTCAGAGGTAAGAAGAGCCTTGGTTTTCTCATCTGTAAAGTGAGATAGAC
ATATGAGAAAGTCAACAGTCATATGAGACAGTTGAAAGATTAACACATTGTA
TCACTTCTCACAGTCATATGAGACCGTTGAAAGATTAATTAATACATTGTATC
GCTTCTCAGCCTTTTGGCTAAGATCAAGTCTAGTAATTAACACATTGTATGTG

GTGATCTGGAAGGGGGCAAGTCGACCTAGTGGCATGGTCTTGTTTAGAAAGC
AGTTCAAGGAGTGCTGTGTATCAGGGCAGAGAGGATTTGACTCAGAACAAGG
GACCAAGGAAGTGAATGTAAAAAAAAAGAAAGAGAAGAGAAAGTTTTAGTAA
TTCTTTGTTGGTTTTTTCTTAAATAGAGACAGGGGTCTCACTACATTGCCCAG
GCTGGTCTTGAACTCCTGGGCTCAAGCGATCCTCCTGCCCAGCCGAGTAGTTC
TTGACTGTGGTAGTAAGGAAAGCTGATCCACGTATCTCTTCTTGAGAAAACTG
TGTATTGTTGACAGTGTGTGTAAATCAGGAAGCAGTGAGAGCATGGAGTTTG
GATTTGGGACAACTGGGTCCCAGTTCTAGCATTTTTCATGTATTAGCCAGTAA
CTGGGCAACTGACTTAACCTTTCAGCCTCAGTTTCCTCATCTTTAAAACAGGC
ATAATAACTAGTTCTGCCTTTTCCCTTAACGGTTGCTAAGAAGACCATTCGAT
ATAAAGCAGGCAAAGTCCCCTGTAACCAATACAGAGGAGTTACAGAAACACT
AAGTATTGTTTCCCTTTGCATTGTGTGATCATGTTCAGCCCTGATACCACAGA
GCTTCTATTCTCCTTTCCTTATTTTGAAGCTCAGGCATTAGAAACATTAGACC
AGAAATTGCGGATTTGTGGGGCCTATAAGCTCAGGTAGCCCACAGATAAGTT
TTGTTTACCAAACATATTCTTCTTCTTCTTTTTTTTTTTAAGAGAGTGTCTCGCT
CCGTTGTCCAGGCTGGAGTACAGTGGCACGATCGTCACTTACTGCCAACCTCA
AGCTCCTTGGCTCAAGCGATCCTCCCACCCCAGCCTCCCTGGTAGCTACAGAT
ACTACAGGTGTGCATCACCATGTCCAGCTAATTTTAAAAACATTTTTAGAGGT
GAGTCTTGCTGTGTTGCCCAGGCTGATCTCGAACTACTGGGCTCAAGTGATCT
TCCTATTCCAGCTTCCCAAAGTGCTGGCATTACAGACATGAGCTGCCATGCCC
AGCATACCAGATGATATTCTTGAAATTTATTTTTATTTTTTATAATCAGATACT
CTCTCAGCAGAATCACAAATGTTTTAATTTGTTAAAAATCTGAAAATTTTAGG
TAAAACTCTAGATTTTCAACTTCTCTTGAAAAGTAAAAAAAAAAGAAAACTGC
AATACTGGGCCCATATTTTGAGAAGCAACAACCAGCTGGAGTTGAGTAGTAG
TGGCTCTTTGATGCCACCACTTTGTCTCTGTGCACACCACTCCTTCCTTTTTGT
CCTACCCCAGGCCCATGTCATGACTTAAGGTGGATACCTGGCCCCTGTGGAA
AGCTCAGTGTGTAGCCTCTGCCTCAGAATATTCCTCAGGCAGAAGGCTGTTCT
CGTCTTTGGTTTTAAACATGCCTCATAGGCAGCAGATTATTTTTCTGTTGCTTC
TGCAGCTGCTTTTATTGTTTAATGCAGTGAGTGACTCAACTTGTTGTTGCTGTT
GTTGTTTCTGTTGTTTGAGACAGACTCTCACTCTGTCTTCCAGGCTGGAGTGC
ACTGGCGTGATCTCGGCTCACTGCAACCTCCACCTCTCAGGTTCAAGTGATTC
TCCTGCCTCAGCCTCCCACATAGCTGGGATTACAGGCACCCGCCACCATACCT
GGCTAATTTTTGTATTTTTAGTAGAGACGGAATTTCGCCATGTTGTCCAGGCT
GGTCTCGAACTCCTGACCTCAAGTGATCCACCTGCCTCGGCCTCCCAAAGTGC
TGGGATTACAGGCATGAGCCACCCCGCCCAGTTGAGTGACTCAACTTTTTATA
AGGGAGTCAGTGCAGTTTTTCAGTTGGTATTCAAATATTTGTAACACCTTCCC
TATCCCTGAACACACACACACACACACACACACACACACACACACACACACCAC
TGTGGTCTGTATTCATCTTGTTTTCCTTCCTCACTTTCGCTCACCATTTGCATTT
CTGTCATGGACTTTAATTTCCTTATTCTTTAAAGTAAGCTATCTCAGAGGATA
ATCTAAATTAACCTGCTTTTAGAACAATTTAAACATCCACATACTTTTACCTA
CCCCTGTTTATGATTTTTATCCTTTCTTTTGATCATTAGCTAAACTGTTGGCAT
CATGTTTAGGAAGGATGAGTAGTCTCACCACTGGGTTGTATCTCCCCTTTATT
TTCTCACCTTTCTCTTGGTTTGGTTTTGGTTTCCATTGTTACAGTGTGATTGCTT
CTTTGAACACAAGGCGCTACATCACAGTACAAAGGAGCTTGGACTCTGCACC
CAGCCCTCCCAGGCTCACACACTTGGGCTGCCACTGCTCTAGGAGCTTCCATT
TACTCATCAATACGGGGGATACTAGTGCCCCTCATGGGGTGGTTATGAGGAG

GCAATGACCTCATACATTGCTTCTCAAGTGTGGTCCCTGACCAGCAGTATCAG
CACCTCATGAGAACTTGGTTAGCACTGTAAATTCTCAGACCCTGCTCCAACCC
TCCTGAATCAAGAACTCTGGGGATGGGGCCCAGCAAACTGCTTTCATAAGCC
TTCCAGGTGATTCTGAGGCAGGCTCTAGTATGGGAATCACTGACTTAACATAC
ACTACAGCACCTAGAACATTGTCCAACACATACCATGTGCTACAGAAAGTGT
TTATTCTTATTACTGTCTAGTCTTTACATAAATGTTTGCATCATCATTATTTAA
TTCTTCTATTCATCTCTCTGATATAGTAGTATGATACTGTTAGCCTTTTTATTTT
TTATTTTTATGGATACATAACAATTATATATATTTATGGGCTACATGTGATATT
TTGATACAAGTATACAATGTGTAATGGTCAAATTAGGGTAATTGGGATATTC
ATCACCTCAAGCTTTTATTATTTTTTTGTTAGAAACAATCCAGCTCCCGTCTTC
TAGTTATTTTGAGATGTGCAATAAATTATTGTTAACTACAGTTGCTCTATTAT
GCTACCAAACACTGGATCTTATTCCTTCAATCTCATTGTATTTTTGTATCCAAT
AACCACCCCCTTTTTATGCTTCCTCCACTACCCTTCCCAGCTTCTGGTAGCCAT
CATTCTACATTCTATCTCCATGAGATCAATTTTTTTAGCTCCCACATATGAGTG
AGAACACTCATATTATTTGTTTTTCTGTGCCTGGCTTACTTCATTTTACATAAC
ATCCTCCAGTTCCATCCATGCTGTTGCAAATGATAGGATTTCATATTTTTTATG
GCTGAATAATATTCCATTGTGTATATTTACTACATTTTCTTTACCCATGCATCC
ATTAATGAACACTTAGATTGAGTCTATGTTGATTATTATGAATAGTACTGCAA
TAAATATTGGAATGCAGATATCTCTTTGATATGCTGATTTCCTTTTCTTTTGAT
ATATACCCAGCAGTGAGATTGCTGGATCATATCATAGGTCTAATTTTAGTTTT
TTGAGGACCCTCTATACTGTTCTCCATAGCCATTGTACTAATTTACATTTCCAC
CAACAACATATGAGAGTTCCCTTTCTCCACATTATCACCAGCACCCATTATTG
CCTGTCTTTTTTATAAAAGTCATTTTAACTGGAGTGAGATGATACCTCATTGT
AGTTGTTTGTGGGGCTTTTTAAAATTTTGTTTTGTTTGTCAGACTTGAGCATTT
CCCAGGCTGGAGTGCAGTGGCATGATCATAACTCACTGCAGCCTTGAACTCC
TAGGCTCAGGCATCCTCCTGCCTCAGCATCCAAAGTAGCTGGGACTACTTGTA
GTTTTGATTTGCATTTCTCTGATGATAAGTGATGTTGAGCACCTTTTTACATGC
CTATTTGCCATTTGTATGTCTTCTTTTCAGAAATGTCTATCCAAATATTTTGCC
CATTTTTTAAATCACATTTATTTTTACTATTGAGCTTCTTCTATATTCTGGTTAT
TAATCCCTTGCCAGATGGGCTTTGAAAATATTTTCTCCCCATGGATTGTTTCTT
CACTTTGTTGGTTGTTTCCTTTGCTGTGCAGAAGCTTTTTTAGTTTGATGTAAT
CTCATTTGTCCATTTTTGCCCTTGGCTGCCTATGCTTTTGAGGTCTTACTCAGG
AAATCTTTGTCCAGACTAATGTCCTTGAGCATTTCCTCAATGTTTTCTTCTAGT
ATTTCATAGTTTGGGGTCTCAGATTTAAGTATTTAAGTCATTTTGATTTGATTT
GATTTTTGTATATGATGAGAGAAAGGAGTCTAGTTTCATTCTTCTGCATGTGG
ATATCCAGTTTTCCCAGCATCATTTATTGAAGACACTGTCCTTTCCCCAATGT
ATGTTTTTGATGCTTTTGTTAAAAAAGAATTGACTGGCTGGGCACAGTGGCTC
TTGCCTGTAACCCAGCACTTTGGGAGGCCGAGGTGGGAGGATCATTTGAGGT
CAGGAGTTTGAGACCAGCCTGGCCAACATAGTGAAACCCCATCTCTACTAAA
AATACAAAAAATTAGCCAGATGTGGTGGCACACGCTTGTAATCCCAGCTATT
CCAGAGGCCGAGGTGAGAGAATCACTGGAATCCAGGAGGCGGAGGTTGCAG
TGAGCCAAAATCATGCCACTGCACTCCAGCTTGGGCAACAAAGTGAGACTCA
TCTCAAAAAAAAAAAAAAAAAAGGAGGGGGGAGTTGACTGTAAACATGTGGA
TTTATTTCTAGGTTCTCCATCTTGTTCCATTGTTTTATGTGTCTATTTTTATGCC
AGTATATTATGTTTTTGGTTACTATAGTTTTGCAATATAATTCGAAGTCAGGT
AATGTGATGCCTCTAGCTTTGTTCTTTATAAGTAGCCTCATTTTAAATGAGGG

ATCTGAGGCTCAGAGAACTGCTAAATGGTAGAAAGAGTTGAAGCTGGGTCTT
CTAACTTCATGTTCAGTGCTCTGTTTCATGTCCCCACACTATCCCACATCTTAA
GAGTGTAAACTAATAGGGGCAAATTTAGATAAATTGGCCAGGCATGGTGGCT
CACGCCTGTAATCCCAGCACTTTGGGAGGCTGAGGTGGGTGGATCACTTGAG
GTCAGAGTTTGAGACCAGCCTGGCCAACATGATGAAACCCTGTCTCTACTAA
AAATACAAAATTAGCCAGGTGTGGTGGCTCATGCCTGTAGTCCCACTTACTCA
GGAGGCTGAGTCAGGAGAATTGCTAGAACCCAGGAAGCAGAGATTGCTGTG
AGCTGAGACCATGCCACTGCACTCCAGCCTGGGAGACAGAGCGAGACTCCGT
CTCAAAAAAAGGAAAAAAAGATAAATAAATGCTTGGCTGTTGTAGATATTTG
TAGATTTCCTTGTCCTCTCTTTTCAGCAGCAGTCCCCAACCTTGTTGGCACCAG
GGGCTGGTTTCGTGGAAGACAATTTTTCCACAGACTGGGGGCTGCGAGTGAG
GGGGTGGTTTCAGGATGAAACTGTTCTGCCTCAGATCATCAAGCATTAGTTAG
ATTCTCATAAGAAGCACTGCAAGCTGGATCCCTGTATGCGCAGTTCACAATG
GGGTTCCACTCCTATGAGAATCTAATGCCGCCGCTGATCTGACAGGAGGCGA
AGCTCAGATTGTAATGCTCTTTTGCCTATCTCTCACCTCCTGTTGCGCAGCCCA
GTTCCTAACAGGCCATGAACCGGTACCGGTCCGCGGCTCAGGGGTTGGGGAC
CCCTTTTCAGGAAACAGCATCCTGATTTTCCTTTGAAGAATCAATCTGCCTTC
ACTCTTAGTGTCTGAGGTTCGAGCTAGGGTCCAGGGAAGGTCTGTGTCCTAG
GCTGCATTTCTTTGACAACATCATTGGCTTAGAGATGGGCAGGTACCCCAAGC
TGGGCCAAGCCCCTGAGCTGAACTTTTACTAGAGCTAAGCCATAGGTAAGAA
GGTGCTTCTTTTTCTTAGATTTGCCAGGATGCTAGGAGCCTGAAGTTTTTGGT
GGTCCTCGTTGTCACTTCATGGAGAGGTCGGCCTGAGAATGAAGCCTATGCA
AAGGAAAACAGAGGGGACAGGGAGAGAGAAATTGATAATAAGTATTGATTA
GCTTATTGGATCCATCCAGTCCTTGGACTTTCAGTAACCTTAGTTCACTAATTC
CCTCTTTTGTGTTTAAGCCAATTTGAGTTGGGTTCTGTCATTTGCAACCAAAA
CAATATAGAATTTATGCCGATAATTTAAGATTTGTATGTTGCTCATAAGATTG
GAAACCTAAGCCTTCCTCCGAACACAAATATGTGAGTAAGAATAAAAAAAGT
CAAATACAAAGTCTTGATGATAACAGTCAATGGGATGGATTTGGAAAGTGTC
TCAGACTAGGAGTCAGAAAACCTGGGTGCTGATCCTAGGAGGTTCACTTAGC
CATGTGACTTTGTAAAATGTGTGAAACTAACGCTGTAAAATATGAAAAATTA
ACCCTGCAACAAAAAAGTAATGTACACATACATATTTTGAATGTTGCATGAT
ATAAAGTCAAGTAAGTAATTTATTAGAGTTGCTCCCAGAACTTTTTCTGTATA
TCCAATAATTAGTTTTCAATAATAGTGAAGGAAAGACCTAGGAGTGGACATT
TCTCAGTCTCTCATACACCATATGCACTGATCTGGGTGACAGAATACCTTTAA
AGATACACTTAAAAATGCATCTTAAGAGAATAAGACAAGCCACAGACTGGG
AAAAATATCTGCAAAACACTTATCTAATAAAGTTTTGTATTCAAAATATGCAA
ACACCCTTAAAACTCCACAATAAAAAACAAATAGCCCAATTAAAAAATGAAT
GAGAGATCTGAACAGACACCTCACTGCAGAATATATACAGTATGAAAAGATG
CTCAACATCATATGTCATTAGAGAATTGAAATTTAAAATAACAAGATACCAC
TACACACCTATTACAGTGGCTCCACTCCAAAATACTGACAACACTAAATGCT
GGAAAGGACGTGGAGCAGCAGAAATTCTCATTCATTGCTGGTGAGAATACAA
AATGGCACAGCCACTTTAGGAGACACCTATTTCTTAGAAAGTGAAACATAGG
CTTACCATATGGTCCAGCGATTGTGCTCTTAAGTACTCATCCAAATGAACTGA
AAACTTATAGCCACACAAAAACCAGCACACAAATATTTATAGCAGCTTTATT
CATAATTGCCAAAAATTGGAAGTGACCAAGATGTTCTTCAATGGATAAACAA
ACCATGGAACATTTAGACAATGGAATATTATTCCAGGATAAAAAGAAACCAA

CTATCAGGCTATTGCAAAGTGAAAGAAACCAATCTGAAAAGGCTACATAGGC
TTCATCTTTAAGACTCCAAATATGACATTCTAGAAAAGTTAAAACTGTAGAG
ACAATAAAAATATCAGTGGTACTAGAGTGAGTGGGGGAAAGAAGGGAGGGA
TCATTAGGTGGAACACAGGGCATTCTTAGGTCAGTGACACTACTTTTATGACA
CTGTAATTATAGATACATGACTAATGCAACATATTACTAATGAAGAAAATTG
TGCGGGGAGGGAAAGTAGGCATATGGAAACTCTCTGTACTTTCTGTTCAATTT
TTCTATAAACCTAAAACCACTATAAAAATAAAGTCTATTAATTTTTTTTAAAA
AATGTGTCTCTCTGTAATTCCTTCTGCCTCTCATCCTGAAGTCCTCTATGAAAT
GGAATCAGGAAAGACAACACCTTACTCTTGAATCCTGGGAGCCCCAAGAAGA
AGTAGGACTGGAATCTCAGGAGTACTAGAAACCAGGAAAATTTGGATCCACT
GACATCCAAAGTGGATCAGAAAGAAAAGTGAGTGCCTCAGACACCAATTGG
AAAAGAAAAATTTATTTACTTTCTCAATGAATATTTATTGAGTGCCTCCTGTG
TAAGGCACTGGTCTAAGAGCTGAAGATACAACAGTAAACAAAGTTTCTCCCG
TCACGAGGCTCAGGTTCTAGTTTGGAGGGACAAAAAGAAAAAAACATGTGA
ATTTATAGACTGTCAGATGCAATAAGTGCCATGGAGATAATATAGCTGGTGA
GTAGTGTGTGGCAGAAGCTGTTAATTGTTCCGTAGAATTAATTGTTCCATTCA
TCTTTTAGAAAAAGAAGCCTCATCTCAAAGTTTACTTGGTTATGCATCTGTAC
ATTACAATCTTCATTTCCCCAGATCCATTCTACTCTCCTGCCCCCTTGCACCTA
GATACAGCCTTGTGACCAAGGTCAGCCCCATGGGAAGTGCTCTGCGTAACTT
CCAGGTCATTTGCTAAAAGATAAAGCTACTTGCCTTGGATTCTTTCTTTCTCTC
TCCTATTGACTGGGAAATAATGATTGAAGAATCCTTGGAAGCCAAAAGTGGA
AGACAGCAGAGCCCCCAGTTGTAGTCTGTTCGATTCTTAGCTGTTACATGAGA
GGAAATTTTTTTTTTTGAGACTTCCCCAATGCCTGGACTTTTCAGTAACATAA
GGCAATATATCCCCTTGTGTGTATAGGCAAGGCTGAGTCAGCTTTCTGTCACT
TGTAACCTAGATTAATTTTTTCTGATTAATTAATTGTTCCAATTAGTACAAGA
ATAATTTATATACTAGACTTTCTGCCTCACCTTGAAACCTGGAAGCTCACTGT
ATCAGTTCTCTCTTGCCACAATGAAGTTGCTTGTAAAACAACCACAGTGGCAA
GCAACAAGTCTACAGTTGGCTGAGTGGCTCTGCTATTCTGTATTGGTCTTGGC
TGATCTTGGCTGGGCTCATTCATGTGTCTGTGGTCAGCTGGCAGACTGGCTGG
GGGCTGGCTAATCTAGCACAGTTTTGGCTTAAATGACCCAACTACCTGGCTCT
GCTTCACAGCATCTCTCATCATCCATTAGGCTATCCTGAGCTTTTTTCATGGTA
AAGCAGGGTTCTGAGACAGACAGGCAGAATGCAAGGTCTAAGCTCATAATG
GTCACAGTACCACTCCCACAGCATTTTGTTGGGTACAGTAAGTCACAAGGAA
GGTCCATATTTAAGGATGGAGGAAATAGTCTCTACTCCGAAGAGCTGCAAAT
TCCTATTGCAAAGGGCATGGCTACTACAGGGAGGGGTGGAGAATTGGAGACA
TTTTTGCAATCAGTCTACCACATTTATCATAATTATTCCCATAATTAAACAGT
GGTGGTAGAGAGAAAAGGAGCAACTGGATACTGGAGCTTAACCAGGTGCCT
ACTAAGCACAAATTCCTCAGTACATGCAGGGGATGAAGCTATCCTATTCTGA
TTCCACCCATTTAGCATATGTGATTTTTTTTTTAAGTTGGAAATAGGCAGAGA
GACCCTGTTTTGAAAGAGAACTTAAGGGTCATCCGGCCCAATCTCCATATTGT
TTCAGAAAAAAGAACCAGGCACCAAAGTTAGGGGACTTGCCCAGGGCCACA
CAGTGAGTGAGGGACACACGTGGGTCTAGATGTTTGCCTGCTGACTCCCCGC
ACAGCAGACCTTCCACGGTCCTGTGCTGTTCTCTTCTCACGCCAAGGCAGAAC
ACAGCAGTGGCCACGTGCTTGCACTTTGGAGGCAGTTAGCCTGGGTTGTCTTT
CTTCCTTCCTTTCTTTTTTTTTGAGACGGAGTTTTGCTCTTGTTGCTCAGGCTG
GAGTGCAATGGTGTGATCTCCGCTCACTGCAACCTCCACCTCCCAGATTCAAG

CTATTCTCCTGCCTCAGCCTCCCGAGTAGCTGGGATTACAGGCATGCACCACC
ACGCCCAGCTAATTTTGTATTTTTAGTAGAGACAGGGTTTCTCCATGTTGGTC
AGGCTGCTCTCGAACTCCTGACCTGAGATGATCCACCCACCTCGGCCTCCCAA
AGTGCTGGGATTACAGGCGTGAGCCACCATGCCCAGCCAGCCTGGGTTTTCA
ACCTGGCCTCTTGCTTATTACTTATGAGACTTTGGACAAATTATTTAACCCCTC
TATCAATGATTGTAATAGTACTGGGCCTCATAGGGTTGTTGGGCAGATTAAAT
GAAAGAAAGGAAATAAAGCACATAGCAAGTGCTCAATAAATTTTAGCTATTA
TATTTTCTCTAAAAATACAGCATTTTCCTATTTGGTTTGTTCTTGTGTGCATTT
AGTCTGGGTTTAGGCATTCAAGAGAGCTGAAAATATCATAATACTAAATATT
TAGATGGCAAAGAATGAATTCAACTTATAAAAGTACCTGGAGTATAAATTCA
CATTTTCTTGTAAGAAGAGATATTTATAATCTGGTTTATTTGTTTACTTACTAA
CAAACATTTACTGAGAGTCTACTGCGAGTCAGGCATTGTAGTAGTTTGCTCTT
GCTGCAACATATTACCACAAACTTAGTGCCTTAAAACAACATATGTTCTGGA
AGTCAGAAGTCTGAAATGGATCATCTGGGCTGTTCCTTATGGAAGCCCCAGG
GGACAATCTGTTTCTTTGGTTTTTCCAGCTTCTGGAGGCTGCCAGAATTCCTTG
GATCATGGCCTGTTTCACTCCAATCTCTGCTTCCACCATCACCTCTTCTCTCCC
TTGACTCCATTGTCTCATTGCCTTCTCTGACACTCTTGCCTGTCCATTATCAGC
CTCTCCATATGGATTGTGCTTTCTTACAGCATGGTGGCCTCAGGGTAGTCAGA
CATGGTGGCTCAAGGCTCCAAAAATGAGTATTTTCAGCAAGCAAAACAAAAG
CTCCATGCTCTTTCATGAATTCACATTAGAAGTCACATAGCTTTGTATTCCAT
GTGGGAGTTGAAGAATGAGAAGTCATGGACACAGGGAGGGGAACAACACAC
TCTGGGTCCTGTTGTGGGATGAGGGATGAGGGGAGGGACAAATACCTAATGC
ATGCAGGGCTTAAAACCTAGATGACGAGGCTAGGAAGAAACTGCATCAACTA
ACGAGCAAAATAACCAGCTAACATCATAATGACAGGACCAAATTCACATATA
ACAATATTAACTTTAAATGTAAATGGGCTAAATGCTCCAATTAAAAGACACA
GACTGGCAAATTGGATAAAGAGTCAAGACCCATCAGTGTGCTGTATTTAGGA
AACCCATCTCACGTGCAGAGACACACATAGGCTCAAAATAAAGGGATGGAG
GAAGATCTACCAAGCAAATGGAAAACAAAAAAAGGCAGGGGTTGCAATCCT
AGTCTCGGATAAAACAGACTTTAAATCAACAAAGATCAAAAGAGACAAAGA
AGGCCACTACATAATGGTAAAGGGATCAATTCAACAAGAAGAGCTAACTATC
CTAAATATATATGCACCCAATACAGGAGCACCCAGATTCATAAAGCAAGTCC
TTAGTGACCTACAAAGAGACTTAGACTCCCACACAATAATAATGGGAGACTT
TAACATCCCACTGTCAACATTAGACAGATCAATGATACAGAAAGTTAAAAAG
GATACCCAGGAATTGAACTCAGCTCTGCACCAAGTGGACCTAAAAGACATCT
ACAGAACTCTCCACCCCAAATCAACAGAATATACATTTTTTTTCAGCACCACA
CCACACCTATTCCAAAATTGACCACATAGTTGGAAGGAAAGCACTCCTCAGC
AAATGTGAAAGAACAGAAATGATAACAAACTGTCTCTCAGACCACAGTGCAA
TCAAACTAGAACTCAGGATTAAGAAACTCACTCAAAACTGCTCAACTACATG
GAAACTGAACAACCTGCTCCTGAATGACTACTGGGTACATAACGAAATGAAG
GCAGATATAAAGATGTTCTTTGAAACCAACGAGAACAAAGACACAACATACC
AGAATCTCTGGGACACATTCAAAGCAGTGTGTAGAGGGAAATTTATAGCACT
AAATGGCCACAAAAGAAAGCAGGAAAGATCCAAAATTGACACCCTAACATC
ACAATTAAAAGAACTAGAAAAGCAAGAGCAAACACATTCAAAAGCTAGCAG
AAGGCAAGAAATAACTAAAATCAGAGCAGAACTGAAGGAAATAGAGACACA
AAAAACCCTTCAAAAAATTAATGAATCCAGGAGCTGGTTTTTTGAAAGGATC
AACAAAATTGATAGAGCGCTAGCAAGACTAATAAAGAAGAAAAGAGAGAAG

AATCAAATAGATGCAATAAAAAATGATAAAGGGGATATCACCACCGATCCCA
CAGAAATACAAACTACCATTGGAGAATACTACAAACATCTCTATGCAAATAA
ACTAGAAAATCTAGAAGAAATGGAAAAATTCCTTGACACATACACTCTCCCA
AGACTAAACCAGGAAGAAGTTGAATCTCTGAATAGACCAATAACAGGAGCT
GAAATTGTGGCAATAATCAATAGCTTACCAACCAAAAAAAGTCCAGGACCAG
ATGGATTCACAGCCGAATTCTACCAGAGGTACAAGGAGGAGATGGTACCATT
CTTCCTGAAACTATTCCAATTAATAGAAAAAGAGGGAATCCTCCCCAACTCA
TTTTATGAGGCCAGCATCATCCTGATACCAAAGCCTGGCAGAGACACAACCA
AAAAAGAGAATTTTAGACCAATATCCTTGATGAACATTGATGCAAAAATCCT
CAATAAAATACTGGCAAACCGAATCCAGCAGCACATCAAAAAGCTTATCCAC
CATGATCAAGTTGGCTTCATCCCTGGGATGCAAGGCTGGTTCAACATACACA
AATCAATAAACGCAATCCATCACATAAACAGAACCAATGACAAAAACCACAT
GATTATCTCAATAGATGCAGAAAAGGCCTTTGACAAAATTCAACAACCCTTC
ATGCTAAAAACTCTCAATAAATTAGGTATTGATGGGACATATCTCAAAATAA
TAAGAGCTGTCTATGACAAACCCACAGCCAATATCATACTGAATGGGCAAAA
ACTGGAAGCATTCCCTTGAAAACTGGCACAAGACAGGGATGCCCTCTCTCAC
CACTCCTATTCAACATAGTGTTGGAAGTTCTGGCCAGGGCAATTAGGCAGGA
GAAGGAAATAAAGGGTATTCAATTAGGAAAAGAGGAAGTCAAATTGTCCCT
GTTTGCAGATGACATGATTGTATATCTAGAAAACCCCATTGTCTCAGCTCAAA
ATCTCCTTAAGCTGATAAGCAACTTCAGCGAAGTCTCAGGATACAAAATCAA
TGTACCAAAATCACAAGCATTCTTATACACCAATAACAGACAAACAGAGAGC
CAAATCATGAGTGAACTCCCATTCACAATTGCTTCAAAGAGAATAAAATACC
TAGGAATCCAACTTACAAGGGACGTGAAGTACCTCTTCAAGGAGAACTACAA
ACCACTGCTCAGTGAAATAAAAGAGGATATAAACAAATGGAAGAACATTCC
ATGCTCATGGGTAGGAAGAATCAATATCGTGAAAATGGCCATACTGCCTAAA
GTAATTTATAGATTCAATGCCATCCCCATCAAGCTACCAATGACTTTCTTCAC
AGAATTGGAAAAAACTACTTTAAAGTTCATATGGAACCAAAAAAGAGCCCGC
ATCGCCAAGGCAATCCTAAGCCAAAAGAACAAAGCTGGAGGCATCACGCTA
CCTGACTTCAAACTATACTACAAGGCTACAGTAACCAAAACAGCATGGTACT
GGTACCAAAACAGAGATATAGATCAATGGAACAGAACAGAGCCCTCAGAAA
TAACGCCTCATATCTACAACTATCTGATCTTTGACAAACCTGAGAAAAACAA
GCAATGGGGAAAGGATTCCCTATTTAATAAATGGTGCTGGGAAAACTGGCTA
GCCATTTGTAGAAAGCTGAAACCGGATCCCTTCCTTACACCTTATACAAAAAT
TAATTCAAGATGGATTAAAGACTTAACATGTTAGACCTAAAACCATAAAAAT
CCTAGAAGAAAACCTAGGCAATACTATCCAGGACATAGGCATGGGCAAGGA
CTTCATGTCTAAAACACCAAAAGCAATGGCAACAAAAGACAAAATTGACAA
ATGGGATCTAATTAAACTAAAGAGCTTCTGCACAGCAAAAGAAACTACCATC
AGAGTGAACAGGCAACCTGCAAAATGGGAGAAAATTTTCGCAACCTACTCAT
CTGACAAAGGGCTAATATCCAGAATCTACAATGAACTCAAACAATTCACAAG
AAAAAAACAAACAACCCCATCAAAAAGTGGGCAAAGGATATGAACAGACAC
TTCTCAAAAGAAGACATTTATGCAGCCAAAAAATACATGAAAAAATGCTCAT
CATCACTGGCCATCAGAGAAATACAAATCAAAACCACAATGAGATACCATCT
CACACCAGTTAGAATGGCCATCATTAAAAAGTCAGGAAACAACAGGTGCTGG
AGAGGATGTGGAGAAATAGGAACACTTTTACACTGTTGGTGGGACTGTAAAC
TAGTTTAACCATTGTGGAAGTCAGTGTGGCGATTCCTCAGGGATCTAGAACTA
GAAATACCATTTGACCCAGCCATCCCATTACTGGGTATATACCCAAAGGATT

ATAAATCATGCTACTATAAAGACACATGCACACGTATGTTTATTGCAGCACTA
TTCACAATAGCAAAGACTATTGTCTTTTGCTATTTGGAACCAACCCAAATGTC
TAACAATGATAGACTGGATTAAGAAAATGTGGCACATATACACCATGGAATA
CTATGCAGCCATAAAAAAATGATGAGTTCATGTCCTTTGTAGGGACATGGATG
AAGCTGGAAATCATCATTCTCAGTAAACTATCACAAGGACAAAAAACCAAAC
GCCGCATATTCTCACTCATAAGTGGGAATTGAACAATGAGAACACATGGACA
CAGGAAGGGGAACATCACACTCCGGGGACTCTTGTGGGGTGGGGGAAGGGG
GGAGAGAGAGCATTAGGAGATATACCTAATGCTAAATGACGAGTTAATGGGT
GCAGCACACCAACATGGCACATGTATGCGTATGTAACAAACCTGCACATTGT
GCGCATGTACCCTAAAACTTAAAGTATAATAAAAAAAAAAAAACCTAGATGAC
GGGTTGAGAGGTGCAGCAAACCACCATGGCACATGTATAGCTATGTAAGAAA
CCTGCATGTTCTTCACATGTATCCCAGATCTTAAAATAAATAAAAAATAAAA
ATAAATAAATAAAAATAAAAATTATATTAAAAAAGAAGTCACATAGCTTTAT
TTCCATACCCCATGGGTTGAAGCAGTCACAGCCCATTCAGATTCCAGGGAAA
GGGACACAGGCCACATCTCTTGATGAAAAGAACATGAAAGAATGTGCAGTTA
TGTTTTAAAAACATCCCAGTAGAGTTCACGAACATGAGTTTTTACAGCAGAC
ACTACATTTCCCTGCCAGTTTACCTGCCTTGGGATGGTGGAGGTCTCTGAAGT
TGGCAGTCGTTTCCTGCAGGATTCTAAGTTGGATGGCAGCAGCTCTCCAGCTC
TGAGGCAACGAAACTGAAAGCTAGTGGAGAGTTGCCTGAATTTTGCCTTCTC
AGGTCTTTCCATAAGTTCTGTGAACACTCAATTTCCTGTATCAAATTCCTTCTT
CTTGAAAATGCTTAGAGTGATACCTGTTTTTTCTACTGGAGTCTGACTGATTC
AAGCTCCAAAGTCTGCCCTCCTAACTGCCTCTCGCGTTGTTCTAAACCTTTCT
GGTGCTCCTGGCCTGCTCCTTTGCAACCCACACACACTCACACATCCAGCATA
CCCTAAGAAGATGACACTGCCTCTTAGTGCTCACAAAAGGAGTGCAAGTTAT
ATGAACCTCAACTATCCTTTCTATCCAACTGGAACTGTATCTGTCTGTTTTTCC
CCCTTCTGCTCCGTCTTAGAAGAAAAGTTCATCAATACTTTTGGGAAAAAGGT
AAACTTTTAAACACGATGCATGGCACCCTTCATTTATCTTTTCAACCTGATTTT
CTGCCATCTTTTTATATGTGCCCATGTTAATTATGGTAGACTAACTGCTTTACC
AAATAGACTCATAAAATTGGTGGAATTATTGCTGCAACAGAAAGACCCAAAG
GTCTACAATGGCCTAAACCCTGTAAAAGTTTATTTCTTGCTCAAATAACAATT
ATAGGCAAGCAAATAATCAGTGGCATATGCCCTTCTCTATGTAGGGACTTAA
ACTAATAGAGAGACAACCATTTCCTTCCCTCCCTCCCTCCCTCTCTCTCTCT
TTCTTTCTTTCTTTTTTTTTTTTTTTTTTTTTTTGATGGAGTCTTGCTCTGTCACC
CAGGCTGGAGTGCAGTGGCGCGATCTAGGCTCACTGCAACCTCCGCCTCCCG
GGTTCAAGCGATTCTCCTGCCTCAGCCTCCTGAGCACCTGGGATTATAGGCGC
CCACCACCACGCTCGACTAATTTTTGTATTTTTAGTAGAGACAGGGTTTCGCC
ACGTTGGCCAGGCTGGTCTCAAAGTCCTAACCTCAGGTGATCAGCCCGCCTTG
GCTTCCCAAAGTGCTGGGATTACAGGCGTGAGCCACTGCGCCTAGCCTGAGA
CAACCATTTTCAACATTTGCTTTCCAAGGTCACCCTTTCCAGACATCCAGAAG
AGAATGTAAACTAAAAATAAAATCCTAAGCCCCCCAACCAACTGAACAGACC
CCCTCTTGGCCAAGAAGACCTCAAGAAAAACTTAAAAACTGAATTCCTGGCC
ATCACAAGAAGGGAAGGTCCAATATGCTTTGTTATCCTCCCTCCCTTTTGGAG
TTTAGGCACAACTGAACAGCATTAGTGTTAAAATAGAGATCGTAAAGCTAAC
AAAATGGACTTATTGTCACAATAAGATGCCAAATTACAAATAGGACCTAACA
CGACACAAGAAGGGTTGAGTCACACATTCTTATAATTCACTGTAACCCAGTG
TACTGGAAAACAATATCTTAATATGAAATATTCCTTTTTTGCTGCCTCCGAAT

TTTTAGACAAAGCTTTATTTCTTTAACCAATTGTAAATTAAACAGTCTCTGAA
TTTACTTATACCCTGTAAGCACCTGCTTCAAGATATCCCACCTTTTCAGGCTG
AATCAGTGTAAATACCTTCCATGTGACTGATTTATATCTTTGCCTGTAACGCC
TGCCTCCCTAAAATGTGTAAAACTGTACTGTAATCCTACTACCAAGGGTGCAC
TTTCTCAGGACCTCTTGAAACTGTGTTCCCCAGGCCATGGTCACTCATATTGA
CTCAGAATAACATCTTTAAAATATTTTAGCCAGGTGTGGTGGCTAACACTCAT
AATCCCAGCACTTTGGGAGACCATGGCTAGAGGTTAGCTTGAGGCCAGGAGT
TTAATGCCCAGCCTGAAGAACATAGCAAGATCCTATCTCTACAAAAAAAAAA
AAAAAATTAGCTGGGTATGGTAGCACATACCTGCAGTCCTAGCTGCTCAGGA
GGCTGAAGTGGGAGGATCGCTTGAGCCTGGAGTTTGAGGCTGCAGTGAGCTA
TGATTATGATCATTCCACTGCACTCCAGCCTGGGTAACGGAGCAAGACCTTGT
TTAAGGAAAAGAAGAAAAGAAAAAGAAAGAAAGAAAAGAAAAGAAAGAGA
GAGAGAAAGAAAGAAAGAGAGAGAGAGAAAGGTAGAAAGAAAGAAAGAAA
GAGAAAGAAAGAAAGAAAAAAAGAAAGAAAAAGAAAAGAAAAAGAAGGA
AAGAAAGAAAAGAAAGAAAGAAAGAAAAAATTTTTTATGGAGTTTAGTTTCT
CCATTAACGGGGGAAGCATACAGAGGGTATATGTGAGAGGCTTTTATGGGTT
AGCCCTGGACATGGTACACATCACTTCTGCTCCCATTCCACTAGCTAGAATTC
AGTCACAAGGCTACATCTAGCTGCAAGGGAGGCTGGGAAACGTCACCTAACT
GTGTGCCCAGGAAGAAGAGAAAATGAGTTTTGGTGAATGACCAGCTAGCAGT
GTGTGCCATTGTGCCTTAAACCTTGGCATGCTTATTTAATTTTCCTATAGTTCA
GCTTTGTCATTTCAAAGTTGAGGATCATCGTGGCACCTCCTTCACAGGGCTAT
TGTAAGGATTAATTACATTAGTACTGTGAAGACCTTAGAACATTGCCTGGCTT
TTTGTAAGCATCCCATAAAAGTTATCTACTATTGTTATTATTCTTGTAATCCTT
AGGGACTCTGCTTTGATACTACTTCCTTACATGAAATCTTTCTTCTCCCTTTCC
CCTCTGGCCCATTCTCTCTAGCACCATCATATTGTATTATAGCTGTCTGTTTTC
TGGTTGGTCCCTCCCCTAAATGCTTCTTGAGGATAAGGATTATGTTTAGCTAC
ATTTTAAGTAACTAACATAGTCTCTGGCATATGATAGGTACTCACTAAATATT
TGTTGAGGGAGTAAAGGAGAAAGTGATAAAGGTGAGAAGGAAAGCCACAAA
GTGGATCAGTATCTTGTCATGGGGAGATCTGTGAAAGGACGGTTTGATAAAT
TGTCCCTTCATGGATTGTAGTCACCATCTTCCAGGTCCCAGAGGGTCACTTGG
TACTGAAGTCTTATCAGCCTAGGGACCAAGGCTGCCTGCCCTTCATGCCCTGT
GTGAGACACCGCTTCTGAACTGCCTCACGTCCTCTCAACCTACCTTTTCTCCA
GTCCCTGCTGCAGCACCCAGCACCCAGAGAAGTGAATCTACTTTCCCTCACCT
TGTAACTGAGTCTTTCAACTTTGAATGCATTTTAAAACTTTTTTCCTCTCTTTC
TAGTCTTAAGATGTAACCTCGAAACTGAGTGTAGAAACTCTCTTTTCCTTAGT
CTTAAAATATACCTTGAAATGTACTTTGAAACTCTGCTTCCCTCTCTTTCCCAC
CAGACTCTCCCTTGCACCATGGACACGTATCTAACTGTATACTTGTTAATAAA
TTCCAGGGGCTGATTTTACACAACAGCCAGGCAAGGAGACCCAGCTGCAGAA
TTCTCCCCTACTTGGAGATTACCTCATGACAGACATCTATAACCTGACTGCAA
CTGAGATGGCACCAGCCAGCACTCCAAGTGGACAACAACTCAAAATAACCCT
TGGAAGAAGACATGCAGGCCTGTATCTTGTGTTACTCTGGCATGGTTCTCATA
CTAAGTATCCCCTTTTTATTTTTTATTTATTTTTAACATATTTTTTGAAACAGA
GTCTCTCTCTGTTGCTCAGGCTGGAGTGCAGTGGTGTGAACGTGGCTCACTGC
AGCCTCAACTCCCAGGCTCAAGCAGTCCTCCCAACTCTTAGCCTCCAGAGTA
GCTGGGACTACAGGTGTGCACCACCACATCTGGCTAATTTTTTTTGTATCTTTT
GTAGAGATGGGTTTTCACTATGTTGCCCAGGCTAGTTTCAAACTCCTCGGTTC

AAGCAATCGGACAGCCTCGGCCTCACAAAGTGCTGGGATTACAGGCATGAGC
CACTGCGCCTGGCCAAGTCTCCCCCTTTAAAATCCCTTCCTTCAGTCTAACAC
TTGAAATGGTCTTTTGGAGGCACAAGCCTGGCCATTTCCCAATTGCTAGCATT
TGAATAAAGTTGCTTTCCTTTTACTCGCTTCTCATATTTTGGCTCTCAAGTCAT
GAGCAGCCAGACTTGCATTCAGTCACAACGTGGCTGCCCTGCATCTCATTCTG
CCCTGGGGTTGCTCTTGCTCAACTTCCTTGGACATTGTTAGAAATGTATTTTG
ACTGTCACATATGCAAGTATGCAAGTGCAAGGGAGTTGGCATCTCATGGGGC
AATGTAACCACTGCAGGACATGAACTAGTTGATAAATATTCGTCTCTTCTGAC
TCTTGAGCAGATGATTCTGAAAGGGTGATGAACTCATCCTGGCTTTCCACGGA
TCTGAGTTTTCTCTGAGTTAGCACTGGAAATCCCATGTTGCTGTTGGTTACCTT
AATTCTGAAATGCATTATGCAGGACTCCTTAGGGGGTCCCAGTGAGATCAAG
TTCCAGTTTCCCACAGCTTAATAACACAGATTGTTTTGCCTTTTCCTGCATTTC
TGGGATTACTTCCCAATATATACTGCCTTCACACAAGCTTGTGTCTTCGGCTC
TGTTTTCTATGGGGAATCTAGCTGGTCTTTGAATTGGGCAATGCTTAGCTTAT
CTGTGCAGTGATGAGTGTTTCCAACTTTAGGAAACTTAAAAAGACAGAGGCA
AGTGAAATAAGTCAGGGCCCCAAACAGCCAAATTATCACTTTCCTGACGTTG
TTACAAAATTAGTCTGGAAGGTATGGCACTGTTCAAAGACAGTGTAATGAAA
TATTTATTATAAGGTTTAATTGCTTGAACAGCCAGATATAGAAAGGACTGCA
GACAGATAGTCAGGAGACTTAACTTCTGGTCCCACTTTTAGTATTTAAAATCC
TATGTGACCTTTGACACGCTTCTCCTTGGGCCTCAGTTTCTTCATCTCCAAATT
CTTGGCTGCTACTCTGCTAAGATCAAGTGTACAAAATTAGGAGGTTAGGCTA
GATTTTGCTTTTACCAGTTGTTCTATTGCAGCAGGTGCAGTTATACATGATGG
GCAAAGGCTTGTGGGGTGTCAGGGGCCTTGTCCCTTCCTAGCGCCACTAGGG
TTAATAGGCCTGGTGGCCTGTTTCATGTCCATTTCACTAGTTGAGCACATATT
GAAGAAAGATTTTATATATGGGTCATAGTGGACCATGAAGAACTCCCAACTT
ATTTCCCTGACAACATTCTTTTTTTTCCTTTGAGACAGCCTTGCTTTGTTGCCC
AGGCTGTAGTGCAGTGGCATGATTTCGGCTCACTGCAACCTCCGCCTCCCTGG
TTCAAACAGTTCTCCTGCTTCAGCCTCCCAAGTAGCTGGGACTACAGGCATGT
GCCACCATGCCTGGCTATTTTTTTTTTTGTATTTTAGTAGAGATGGGGTTTCAC
CATGTTGACCAGGCTGGTCTCGAACTCCTGACCTCAAGCGATCCACCCACCTC
CGCCTCCCAGAGTGCTGAGATTACAGGCATGAACCACTGCACCCAGACCCCT
GACAAGATTCTCGAGATTTAACAATCTCAAATCTTCTTTCCGCAGTGCCAACT
AGACAAGATATCAGGGCCAAAAATATTTGGCTCTGAGATGTAAGAAAGGTCT
TTCTAATTACTGTAGAACTTGAGGTCATGATCATCACAGCCAGTGGTGTGCTG
CTCCTACCTGCTTGTAAAAGCCAGTTGCTAAATATTCAGGAATGTTTCAAGCC
AAACAACCACTGGTAGCTTGAAATCAGTCATGGGGAGAGTATTGACACCACA
GAAGTTGACAGATGCTACAAATCCATGTTCTCTCCACATCCACCCCTCCCTGA
GCCAGTTTACCAGCAATTCCCCGCATCCAGCTCTTAACACTGGAAGTTTTTAT
CCAGGACTATGGCTAAATTAAATATGTGTGTCTTAAAAAGAATTTTTTTCCCT
GCTTTTTGGTTTGTGGATACAAAGAGCTGAATCATTAAAACTTGTTTTTCTCTC
CACCTCAACAGTGGCTTTCTTTATAGATTCTATGAATCTGAAAGATACAAGTT
CCACCAGAGGTCTTGGTAGGGTTCCACAGCCCCAAGCTCTGCAGGTTCTGAC
AAAGGATAGGCTGGGCCCTGCCAGGAAAACTCACAGAACCACATCAAGGTG
TTTGTGCTGCTGCTTTAAGAGTTTGTCCTCAAGACAAATAAGAGTTTGCCTAT
TTGGGGGCTTTGGAAATTTTTTTTTAGGATTTTGGAAGCCTGTTCTGTTACTTG
CTAGTTGTGTTACTGGACAAGTTAGTTAACCTTGCTGATCCTGGGAAGTAATC

TCAGGAAAGCAGGGAAAGGCAAAACAGCGTGTGTTATTAAGCTGTGGGAAA
CTGGGACTTGATGCCCCTGGGACTCTCTAAGCAGTCCTGCAAATTGCATTTCG
GAATTAGGATAACCAACAGCAACACGGGATTTCCATGCTAAACCAGGGAAA
ACTCAGGTCCCTGGAAAGCCAGGATGAGTTCGTCACCCATTCAGAATCATCT
GCCTAAGAGTCAGAAGAGGGGAACATTTATCCACTAGTTCATGCCCTGTAGT
GGTTAAGTTGCCCCATGAGATGCCAACTCCCTTGCACTTGCATACTTGTATAT
GTTCAAGTATGGTGCACATGCTTGAGACAGGATCTCGCTCTGTTGCCCAGGCT
GGAGTGCTCTGGCTTATCACGGCTCACTGCAGCCTCAACCACCCGGGCTCAA
GTGATCCTTCCACCTCAGCCTCCTGAGTAGCTGGGACCACAGGCATGCATCAC
CATGCCTGGCTAACTTAAAAAATTTTTTTGTAGAGATGGAGTCTCACTATGTT
GCCCAGCCTGGCCTCAAACTCCTGAACTCAAATGATTCTCCTACGTTGGCCTC
CCAAAGTGCTGGGATTGCAGGAATAAGCCACTGTGCCCAGCCAACCTTGCTG
AGCTTCACTTATATGTGGAATGGATGAGACCACTTCAGAGGGTTTATGGGTTT
GTCATGAAGCTCAGGGAGCTTAAGGCTCAAGACACCTCACTTGTACAAGCTT
CTCACAGGAGGGCTTTAGGTACACTGTATTTGTAATTTTCTAGTCTAAAGACC
CCTGCCCTCTTCCCAAAATGTGTCATCTTCAGAGCTGCACAAACCTTGGATCT
GATCCTGCTTGTGAGGATGAATCACGTGTAAAACACTTATTCTTGAGTCTGGC
CCATTGTAAAAGCCTCAGAAATAGTCCTTGCTGTTGTCAGTGGAAGTACTTTC
CCTCCATTATCACATATCCACATATTCTTTCTCCCACTGCCTAAATGTCAATGC
CCATCCACGTCTGCTCAGAGTTTGCTTCTTTTCCTATAATGTCAGTTTCTTGCC
AATCACTCCCCACTTGCCATTTTGCCAGTCTTATTCCATAAGGTGCATCTGCA
CTTCCTCACTGCGTCCTTTTTCATAATGGGAAAACAATGTCCAGCGATTATAT
TCTTCACAGAAGTGATGTGACAGCAAAGAGAGATCATAAATGAATCGCCTGA
AGCTTGTTGGTGTGAGGGGGAAACCAAAGTCATATTAATACTTAACAAAACA
AGCAACAGGACAGGAAAACAAAAGGTAATTAAGGCAAAGCTGTGATGTTTT
GCCAGTTGTTAACATAAGAGGCCAATTGTCAGCTGACTGTGATGTAAAGACG
CTTCCTTTAAGAGCATGTGTAATATGTATCTCAGAATCATCAAAGGGCTCTAA
GTCACCCTAATAATGGGTCTGTACCACAAACAGAGAGAATGCAAACCACATT
TTGTCTTAAAAGACACAGCAAATTGCACTGCAGCTGTAACAAGAATCTCAGA
GTCATTTGCATTAACTGGGGATGAGTGAAGGGGCTAAGTGGAGTGTCTGTGT
GCAACATGGCACTTCTTCCTTGACCTGTGAGAAAGGAACTTGACAGCCAGGC
TCAGTGGCTCATGCCTGTAATCCCAGCACTCTGGGAGGCTGAGGCAGGTGGA
TCACGAGGTCAAGAGTTCAAGATCAGCCTGGCCAACACAGTGAAACCCCGTT
TCTACTAAAAGTAAAAGAAAAAAAAATTAGCCGGGCATGGTGGCGGACGCC
TGTAGTTCCAGCTACTTGGGAGGCTGAGGCAGGAGAATGGCATGAACCCGGG
AGGCAGAGGTTGCAGCGAGCTGAGATCGTGCCACTGCACTCCAGCCTGGGAG
ACACAGCGAAACTCTGTCTCAAAAAACAAAAAAGGAAAGAAAAGAAAGGAA
CTTGACTTATATACACTTAGGTGCAGCCATCATTGAGGGCTTGTTGTGCAAGG
TGCTATGGATGGTGATGAGTAAAACCAGGTGTCCTTTCTATTATGCTGTGCTG
GAAGCTTCTTTGGCAAGTGAAGGGAGTGTGGCCTTTGGGATCAGATGAATCT
GGTGGAATCCTGGCTCTGTGCTCAGCATATGATTTAGACATTTATGTAACCTT
CTTGAGCCTCAAGTTTCCTCATCTGTAAAATGGTAACAATACTACCCATCTTA
CAGAATCACAGAGAGGATTAAATGGGAAAAAAAAAGACAAAGTGTCTGAAAT
ATAGCAAGTTCTTAATAAATATTAACTTTCTTACCCCCTTCTGGAGGCATAGA
ATCTTAGTGCAATCTTGGTACTCTCAGAAACTGTTTATGTAGCTCATCTGAAT
CTTATTTTTTTAGTAGTTGAAACATTTTGTCTTAATACAGAAGCAATACATATT

TGTAGCACAAGAATTAGAATATTTATTTATATATATATTTATTTATTTTTGAGA
CCAAGTCTCACTCTGTTGCACAGGCTGGAGTTCAGTGGTGCAATCTCAGCTCA
CTGCAACCTCCACCTCCTGGGTTCAAGGGATTCTCCTGCCTCAGCCTCCTGAA
TAGCTGAGATTACAGGCGTGCACCACCATGCCCGGCTAATTTTGTATATTTTT
AATAGAGCCAGGGTTTCACCATGTTGGCCAGGCTGGTCTTGAACTCCTGACCT
CAAGTGATCCACCCGCCTCAGCCTCCCAAAGTGCTGGGATTACAGGGGTGAG
CCACCGCGCCCGGCAAGAATTAGAATATTTAGATAAGAAAAAAATTCAAAAT
TACAAATAATTCTAATACGTTAACACCCTTGGGTGTATCCTTCCAGGACACTT
TTATGCCTGTTAATACATACGTATTATTAAAAATATAATTATACAATACATAT
TATTTTATAACCTGCTTTGTTTTCATACGTTACGTCATTATGGGAAATCTTTGA
AAAAATCTCTCCCATGAAAAGGCCAGCTAACAATTATGGGAAGTATGGGAAG
TGGTTTCGATTAACATACTGTGACAACTATCTATTATTAAGAAAGTCACGACA
AAATTTTGGTGCCCCTTTCCCAGATGAAGGCCACATAGCTGCTATAAGGGAC
AAGGACCAGTAATATCTTCAAATCCTCTTTTTGAAGTGCCTTCATTTAAAAAT
CATGCATTTTCTTTTAAGAGTTTTATAAAATCTAGGAGGAAACTTATCCCTTT
GTGTTTAGAACAAGGCAAAAGAAATTTCTGATTGGATATTGTTATGGGCCCA
TGACTTCTGTTGCAGAGAAGGTAATAGAAAAAGGTAAAATACTTCTACGCTC
TATAATTCACCTTGCTGGAAAAAAAACAACTGGATTGGCTTGACAGGGGCTT
AGACGGGTGACCAGGTTACTGTTTGGTTGGTTGAGAGACGGAAGCAGTACAG
AATGACAAAAGTGTGTGGTGGGCCACCGGCCACTGGTTCATCATAGCAGGAC
CTCAAACCAATGTCTAGTCCATGAATGTTTATATATGGGTTGGTATATGAAGG
TGGATATTTGCAAACAAATGCTTAGTTTTAGTGTCAGGATTTTCTTCCTAATT
AATGAAAAGAGACTGTATGTTTTCAAGTTCTGTAGGCCTAACTAGAAAGAAA
AGAGTTCAGGATTTCAGATTGTGCTACTTTCACAAGATGTAGGTATATCTTTA
CCAAAACACACATGGCTATGCACATGTTCAAGTATCTTGTTATAAGAAGGGT
GTGGTGTAAGTGGAAAAATTGCTTCTGTTATTCTTGTGAGGCAGTGTAACTTA
GTGGTTAGAAGCACTGATTGAAGAGCAAGACTACCTAGTCTTGAATTCAGCT
TCACCAACTGTTAGCTGGGCAATCTTGGGCAAGTTAGTTACTCTTTCTGAGTC
TCTATTTTTCTTGACCTGTAAGAAAGGAAATTGATGGCCAGGAGTGGTGGCTC
ATGCCTGTAATCCCAGCACTTTGGGAGGTTTAGGCAGGCGGATCACAAGTTC
AAGATCAGCCTGGCCAACACAGTGAAACCCCATCTCTACTGAAAGTAAAAGT
AAAAAACAAAAAAAAAACAAAAAAAGAAAAAACAAATTAGCTGGGCATGATG
GTGGGTGCTGGGAGGCTGAGGCAGGAGAATGGCGTGAACCCGGGAGGCAGA
GGTTGCAGTGAGCCGAGATCGTGCCACTGTCTTCTAGCTTTGAGACAGCGAA
ACTCTGTCTCAAAAAAAAAAAAAAAAGAAAAAGAAAAGAAAGGAACATGACT
TATGTATATTTAGGTGCAGCCATGATTGAGGGCTTGTTGTGCAAGCTGCTATG
GATGGTCATGAGTAAAACTAGGTCTCCTTACTATTATGCTGTGCTGGAAGCTT
CCTTATGGGTATAATAATAGTTGCTCCTATTAATGAGTATATCCACCTCATAG
GGTTGTTGAAGGATTTGAAAAATATTTTAGAAAGTAACAACTTTCAGCATTTA
GAACAGTGCCTAGTATATAATAGGAAGTATGTGTTAGCTATTGCCATTTTATT
AGAGTTTTAACAGGTCAGTCCAACAGAACTGGCAATTTCCTGAGTGATACTTT
TTTTTTTTCCTGGGACAGAGTGTTGCTCTGTTGCCCAGGCTGGAGTGTAGTGG
TGTGATCTTGGCTAACTGCAACTTCTTCCTCCTGGGTTCAAGCCATTCTCCTGC
CTCAGCCTCCAGGGTAGCTGGAATTATGAGAACGCACCAACACGCCCAGCTA
ATTATTGTATTTTAGTAGAGACGGGGGTTTCACTGTGTTGTCCAGGCTGGTCT
CCAACTCCTGACCTCAAGTGATCTGCCCACCTCAGCCTCCCACAGTGCTGGGA

TTACAGGCGTGAGCTACTTTTAATTCATTCATTACTCAAACAACACGGTATCC
TGGATGTCTATATATGGTTAGGCACTATGCTAGGCTCTGGGCAACTGAAAAA
AAAATGCCAATATTCTATATCTTTGAAAAGTAAAAACTGACTACTCGTGTCTT
CTCAAGAAGCTTTGTGACTGAGGTTATGCAGGTCTTTTTATCTAGAGTAGGCT
TCCTGGAAGGAGGACAAGACTAGCTCACCTGCAAAGAACTCAGTCACTTATC
AGGAATGAATGAAGTGCAGAGCGTATTGGGATTCCCTAACACCAACTCTTCC
TGAACATGCACCTTTGTCAAACCTGCCACTGTCAGAGCTGCCGACACAGGCA
ATGGATGGGGACATCAGGACAGGGCTTGGGGTGGGAGCCTGCAGTTCCTGGA
ATTTGCCCTGCTGACCTCACACTAGGTGATTTTATCCTACTTCCCAGAAACTTT
CCCCCAGTTAGTCTAAAGTTGGGATGAAGCTACAAATTATTTGCTCTGAAAAT
TCTGAGTCTATTCTAGAGTTAATTTCCTGTACATGTTACATGGTTTGCATTATT
AGAAGGGCCAAGGGGCCCCGGCAGTGCATTCCTTTTCATTTTCTGTAAAAGG
GATCTGTGGGACTCTTTCATCTTCCATTAATGATGAGAATGTGGAAAGGGGA
GGGTGGGGTAAGAGAACTAACAATTATTGAGCACTTACTCTGTGCCAGGTAC
TTTGTACATGTTCCTGGCACAGAGTCAGTGCTCAATAATTATCGGTTTTTGGG
GGAACTGAAACGAAAATCTGAGAGGCCAGGGGCCATGTTTTTGTCAATACGT
CTATTTTGGCAATGAGCACATTACCATGTATATTGTCAGTGGTCCATGTTGTT
GAATGATATGATTATACACATGTTATGTGTGCATATCCACCACCTATCTATGT
ATCATCACTTGTCTATCACTTATTGAATCTTAGCTAAAGCTCATGCTATTTAA
AATTACAATTGCCTTGGTCTCAGTTGAATACCATCCCATAATTTCTATGTGAA
ACAATTGATTTTGTGATTTCTATTTCATAATTGCAGAATTAACTAATTTATTAT
TCAAGTCCTTTGAACTTAAAGATGTTTTCAGTAGTGTGGATTGAATATATCTA
TGGGTTTGATGAAATTAACTTTTTTTTTTTTTATGAGACGGAGTTTCACTCTTG
TTTCCCAGGCTGGAGTGCAGTGGCGCAATCTTGGCTCACTGCAACCTCTGCCT
CCTGGGTTCGAGCAATTCTCCTGCCTCAGCCTCCTAAGTAGCTGGGATTACAG
GCGCCCGCCACCATGCTCGGCTAATTTTTTGTATTTTTAGTAGAGACGGGATT
TCATCATGTTGGCCAGGCTGGTCTCGAAGTCCCGACCTCAGGTGATCCCCCTA
CCTCGGCCTCCCAAAGTGCAGGGATTACAGGCATAAGCCACTGTGCTCGGCC
GAAAGTAACTTTTAATTGTGATAATTACATTGCCCTTATTTCCATTACGCATG
AAACAGAACTCCTCTGTCTTCTTATGGTAAAATTTTGGCACTGAAGACCAGTT
GAAGATGAAGGATGCATCTTGGTAAGTTGAAAGAGGAGAGAGAGTGGGGCC
AGGGAAGCCACAGTGGGCAACAGATTGGCATCCCAGCTTTCACCCCTCTCTA
TCATTCTTAGGTATTTCAGGTCAGAATTAAGAGGAACTAATTGGAGATATCAT
CTTCTTGGAATTGTTTGTGCCTCATAACACTTTAAAAATGCATTGACGCAGTG
ATTTCTGAAGTTCAGTCATTTGTGTATTGCCTTCACGAATTTTGCTATACCTTT
GAATCACCTGTACTCATACTGACTTGAGTTTTTAATCATCTCACTTTCTTAAAA
TTAAATACTTTTCTTTTTTGGGAAGAGGATCTCATATTATTACTTTTTAAAAAA
TCTCTTTTTCCACACGTTCTAGTTATATCATTACTATAAAATTGAAGACAGGC
TTCACTTGTTAAAAACTGAATGTAACTATAAAATTCAATCCAAGAAAAACAA
AGAAAATGTTAGACAAAATTTTAGCTAAGCCTTATTTTTCTTGGACTTGAAGC
CTAAGTTTTGCTCCCTGTCAGAGGGTTGCTAAAAGACCAGCAGCAAACTGAG
AGTTTACCTTGGAGGTAATCAGAAAGAAAAAATAATTTTAAGGGGAATAACT
TTCTCCCCCATATAATTCGGTGTATTTTAATGCCTGGTTGAGCCACTGTAATA
CTATGGAAGATTATGTCATCTTCCGTAATACACATTATCTCACACATGGCAAA
ACACTTCATGAGAAGAGAAACTGACATCTCATGATTTTGACCTTCACCACATA
TACAAGTTTTTTTGTTAGAAATATCACTCACGAAAAATGGATGAAGTAGTCTT

CCATGCAGAAGTTTGGACATTTCAATTTAATGTCTCTGGTAAAGATTTTCCAG
TTAAAAGCACATTGACCACAATGTTTCGTCTTCTCTTTAACACTGAAAAGGGG
AGGAACGGCCTTATTAGCAGTCAATACTATCAAAGTCAATGGAGGCAAGGGA
CCAATGGCCCATCATAGTCCTAATATCACTTACTGTTTTCGAAGGAGGACATC
TTTCATAGTGGTATCTACCATCTCTTGGGGAGAAGAGCAGGAATGGATAGAT
TAACCTCTTCCAGACATCTGCTCACAGTCCCGATGCGGTTTCTGACTTAGAGG
GTTTTTCTTAGAGATCTTCTCTGCTGTCCTCTGCAGCTGTCAGGGCATTCTCCA
GATGGGGCCTGGTAGGAGTCCTTGAATTGACTCAGGTCCCACATCTCCCTGCA
GTTCATTTATGCTTCAGGTCAAGGGTCACAAATATTCACTTGATAAGGGATGA
CAGATTAGTCACAGCTAGCTGCTGAAAGTGGAGTTGCATGAAGTGCACATTT
AGCTTGCATGAGGGAGAGCACAAATTGGAAACTTATCAAAATTGCCTTGGTG
GGTCTCTTTCAAGGCTTCTTTGGAGGCTGCTCTCAGATCTATTACCTCTGGAA
ATTCTTGAGGACTTTTAAAAATACAATGACATCTCATCTCTTCTTAAATTCTGT
CTAGTGGACAATTTGTGAGGGGTGTATGTGGGAATCTGCAATTGAATTCTTCC
TCAGGTACGGTGAGTTGGGCCTCATTCCAGCTCCACCGCCATCATCATCAGAA
TCCTAGTGGGAATTTCCCTTCCTCAGAAGACTGAAACTCCACCTGTGCTGGTA
ATGCCAGGCCCCCTACATAGCTCCAGTCTATTACCTCAAGGGAACACAAAGG
CACACTCTTACAATATATATTTACACCTGAGCTTGTAAATATGACAGCTAAGA
ATCAGAAATGTTGTTTTACTTAAATGGGCTTGTCAGTAGATTGGGAGTATATT
GTATTCCAAGGAAATGCTTTTTAGATGTAATTTTAGGTACAAACTCTGGATTT
TCCTTCACGTACAAAAATTAAAACTTTGATTACACACTAGCAGAAATGACAG
CAGTAAGAGTTTTTTTATTTTAATTGTATTTATTAAACAATTAGGAAACACTG
TATACTTTCTAATTGTTCTAAGTTGCTTTATAATTCTTAATTCATTTCAACTCT
ATTACAACCTTATGATCACCTACTATGTGACTGTAAAGCACTGTACTTTATTA
CATTAGCTCTAATTCTTACAATCATCCTTGAAGGTAGTTAGTACTCCCATTGTT
ATGAATGACCACTCTTCAGAAGCAGGAGGGACCCTCATCCAAATTTGATTTG
GATGTCAAAACTGATGATGCCACACATCCACCTCACATAGGTATGAAAACTA
TTATTCACATAATGAGGCTTTCTGGGGATAGCAGTGTGGTTTCCAAGCAGATA
AAAAAAAATGGCTTGAGAGAGAGCACAGAAAGGAGACTGGCTTAGGGTTTG
TTTTGTTGGTGGTTGAGGTGGGGCCAGGGTGAGGGTTCCCACACATGGTTTGA
AGTTGCCATGCAGCTTCCTAGTAGCAAAGGAGGGAATATCCTGGTTTTCTTAT
CAGTTCTCCTGGATATGGGCAGATGAGGAAGAGGGAGGGGTAAGGCTTAAA
AGCTGGCAGTAGTCAAAAGTCAAAAAATGGAGTCAGATTCCTTATTATACCC
ATTTTACCAATGAAGAAACTGCGATTCTAGGATATCATGCATTGTGTCAGGCA
TGACTTTAGTGCAGTAATTCTCTGTGAAACTTAGCCCCACACTGCACTGTACT
GTGGTCTTGGAGAAGTAGAAGCAATAGGAAATACACACACACACACACACA
CACACACACACACACACACACACACACACAGTGATTTACTATAAGGAATTAG
CTCACATGATTAAGGAGCCTGAGAAGTCCAAGATCTGCATTCAGCAAGACAG
AGACCCAGGGGAGTTAATGGTATAAGTTCCAGTCAGAGTCTAAAGGCAGGAG
AAAATTGATGATGTCTCAGCTTGAAGACAGTCAGGCAAAGAGAAATAATTCT
TTCTTACTCAATCTTTTAATCTATTCGGGCCTTCAATGAATTGGATGAGGCCC
ACCCACATTGGGGATGGCCATCTGCTTTACTCAGTCTACCTATTCAAGTGTTA
ATCTCATCCAGAAACACCCTCACAGACACACTCAGCAGTAATATTCAGCCAA
ATATCTGGCTATTTCATGACCCATACAAGTTGATATATGCAGATAACCATCAC
AGTATCCACCTTGAAAATGCACATCCCTTATTATTGAAGGGGAGTGGGAGGC
AGAGGAAATCCTAAACACCATTGCAAATCTATATATTCTAGAGAGACTATGA

AAGCAATGTAGCTTGGCATGGTGGAAGGAGCATGGTCTTTGGGATCAGAAGA
TTTGGTTATAATTTGGCTTTGCCATTTATTGGCTGTGTTATCTTGAAAATTGCT
TAGCCTTACTGAGTTTTAGTGATACAGAACATCTCCAGTGACATGCAAATTTA
TAAAACATCCCATTCTATTTTGGGGGTCTCCATTGGAAAGCTCTTTCTTAAAT
AATGGCATTTCCCTATATTAGGTTTGGGGTGCATACATAGTCTCTACTGGTAT
ATCAAGGCAAGTTACCAGGAATTCCAATGTATAAGGACACAGCTGGCCATCT
GGAACAAATATTGGAAGGGATTTTGGAGAACAGAGGTTAGATCCAGGGGCA
ATAAGGTCCAGACTCTTCACAAGAGATGAGATGAAGGCCATGCTCTGCCTAT
TTAGGAATCCACAGGACACTGAGAGTTACCCCAAGACAAGAGAAAGTTCAA
ATACCCAAGACATTGATTGACTACTGGCAATAGTTTTTGGCACAACCCTGGG
ATTTGCTGCCCAAATGTTGTTTCAGTCTTGTCCCAGGAGAGCTCAGTTCTCAG
GCTGGCTCCACAGCCTACCATGTTAGCAAGCCCAAAAGTGAATATGTCTTTTG
TAATTTTTCCAGCAAAAATTCCAGGGCTGACTCTCATTGATCCAAATTTGGTC
ACAGGCCCATCCATGAATCAATTGCTGTGACCAATTTACTAGGCCTGAGTCAC
TGGTCTGTCCACCCCAGGTGTCTTGTGGTCAGCTCACCTGAACAGCACGGACT
GAGACCCAAGAAAAACTGATGCACTGTTACTAAAAAAGTGAGGAAGTGAGG
CTGAGTAGGCAAAACAGCAGATGTCCATTACAGGAAAGATCTCCAAAATGTA
ACCTCACTCATTTTTTCTGTATGTGTAACTCTGTGTGAACTTAGCGCCTGCCCC
AAGGGTGGCATTTACTTAGACTGTGATGAAAATAGAGACCCTGGATTTGTGC
AGTGCTCTGTCTTCCCCCCTCTCTTTCTGCTACCACGATTTCTCCAACTTCTGG
TTGTATCAGGTTTCAGCTAAGGGAGGGAAGACATGACTTAACGCATAGGCTT
CCATAGAGATTCCATACTGGGACTTCAAATAGGTCACACTACTGGTGAGTCC
CAGCACACAGATCCTAGTGCCAAGATACTCATGCTGTTGATTCAGAACTTCCA
GACATACATGTGCTTCCTCTGACAGGGAAGGCTGCTACTACCTATATCTTCTG
AATTGGTTCATATTAATCATAGGTCATGTGTACCACTGGTAACAATAGTCATC
GTTTAGTGAATGTTTACAAGTGAGAACACTGCCTATTCATAAGCTTGAAATTA
TCTGTGAATTTGGGAATGTGTGCCAGCCGTAAGCTGACCAGACATACTTACTG
GTCATTAATCAACAGGGTTTTGTTTCCTTATCCCTATGTTTGACTGAGACAAA
TTCCTCTCCCTACATCACTCAAATGTGGATACAGAAGTCTTCTCCCCTTCCTAT
TTATAACCTCAAAAGGTTGCAATTTGGACTGGAGGTAAGGAGGAGATAAAGT
ACTTACAAACTACTTGTGGAATTCCCCCAGTCCCCTCCACCCCCTGCTTTTTCC
CTGTGTCTTGACCAAAAATCACAGAGTACCTTGATCACACTATGACACAACC
AGCAGCAGGCTTTCCCAGCAGGCTTGACACCAGGGCTTTGAACATTCCCAG
GCCCTCATACAGGTATCAAGGTTTACGAGGAAGAAACTGGTCCTAGCCTTAG
CCCAAATCCTTAAACCTTTATATAAACTCCATGCCCTGACCTCCTCACAGCAG
ACATAACTAGATAGAACACCGTTGTCTCTTGCTGTTTGTTGCAAAGATTGCTA
CAGCCTTCTCTGTGCCTAAGTTTTTCTAATCAATGCTTTGGATGGATCAAAAA
GAAAAAAAAGAATTTATAACTAAAAGGAAAATATTGTGTACTATATATTATA
TATAGCATATATAATATATAATCTGTATAAAATACATGTAACATATAATCCAT
TGTATGTTATATGTAATTTTTATGGAAATACAACAAATTATAAGTATAATAAT
TATGTATGTTACATATATATATATTTTTTCACGTTTTTAACTTGAGGTTTAAGT
ACCTGTGATCTTTTTTTTTTTTTTTTTTTTTTGAGACAGAATCTAGCTCTGTCAT
CCAGGCTGCAGGGCAGTGGCTTGATCTCGGCTCACTGCAAGCTCCACCCCCT
GGGTTCACGCCATTCTCCTGCCTCAGCCTCCAGAATAGCTGGGACTCTAGGCG
CCCGCCACCACGCCCGGCTAATTTTTTGTGTTTTTAGTAGAGATGGGGTTTCA
CCATGTTAGCCAGGATGGTCTCGATCTCCTGACCTTGTGATCCGCCTGCCTCG

GCCTCCCATAGTGCTGGGATTACAGGCATGAGCCACCGCGCCCAGCCTATCT
GTGATCTTAACCAATTTATCTTGCCTTTTCTGATGCATGACTAAGAACTTATG
GGGAATTCAAATAGATCACTAATATAAAGACAATTTAAAGTGCTTGACTTGC
CAGTGTTCTTATAGCAAAACTTTCATAAAGTTTATGACACTTCTTTGATTTCTA
AGGTTTTAGGGGTACAGGCATATCTTATTATAGAGTATTTTTCCTCTTTCATTT
AAAGATTTAATTTACTACGCATCATTTCCAAGCTCCTTGACCTTTTCTGGATCC
CCACCACCACCTGCCCTAATGGATACTGACACCAGCTGTGAATATGTATAAT
ATAAACTGGCCAAAACACATGGCCTGGGCAGGTTAATAACTTGCTGCCCCTT
CAAACTGATATTATTTAAATTTGTATTTAACACTTGAACACTGTCATAGGCTG
ATCCTAAAATGAACCCCAATAATCCATGCCTCTTGGTAGTTATGCCCTGTGTA
ATCTCCTCCCCATGAGTGTGGGCTGACCTAGCAACTTGCTTTTAACTACTATA
AAATAGCAACAGTGATGGGCTGTCATTTCTGTGATGGTGTTACATAAGAGTTT
TACTTCTGTTTTATTAACAGACCCTCTCAATGCCTTCTCAGCTTGCATACTTTG
ATGAAACAAGCAGCTGGGTTAAGAGATGTCCATGTCGCAAGGAAATGAGGG
CAGCCTCCATCAACAGCCAGCAAACAACTGAGGCTCTCAGTCTGACAGCCCA
TGAGTAACTGAATCCTGCCAACAAGCATGCAAGCTTGGAAGCAGATCCTTCC
CCAGTCAAGCTTTTGAATGCAACCTCAGCCCCTGCTGACACTTAATGGTGCCT
TGTGAGGGCCCTGTAGGCAGTGGAGCAAGCTAAGCTGTACCTGGATTCATGA
CCCACACAAACTGGGAGGCAACAAATGTGTGTTGTTTTACACCACAAAATTT
GTGGTAATTTGTTACACAGTGACAGATTAATACAAGTACTGAGTGGGGAAGG
TTGCATATACCATTCAGCCAAAGCTCTCTTGTTAACTGGAACCACCCTAATTA
CAAGATAATTTAATGAATGACTGTGTTACCTGAACACACCCTTCAGAGACCC
CATCACTTGCATGAGTCAGAGCTCTATGGGCTTCAAGTGACAGAAATCCATTC
TAATTAGCTTAAGTAAAAAAGGGATTTGTTGATTTACATACCTGAGAGGCCA
AGGGTGTACCAAGCACCCTCAATTTCAGATCTACTAAAAAGGACTCTGACTG
GCCCTTTTTAGATACATGCCAATCATTAAAGATTTAATGTATCTATAATAAGA
TACTATATAATAAGATATTAGGACCAATCATTGTGTCTAGTTGGATAGAGTTC
TCTGAGTGGTTGGCTTGGGTTTCATGCCTACCCTTGCGGTGGAGAAGGGGAG
AAGAACATTACATTTGAAAGCCTTATCATGTAATATACAGATAGTTACCCAA
AGAAATTTTTTTTTTTTTTTGCTCTAAAAAGGCGAGAATGTACACAGGGCAGG
CAGGCAAAACAACAGCTGTCTACTATACTGCCTACTACCCAGCTGGGAGGCA
AAGTGACTCCATCTTGGATGCTAACCTGCCATGTTGACTTCTGATTAGCCACA
ATCCTGTGAATATCTCCTGATTCCTACTTTATTTACTGTTTGTGTATAAGAACA
TGTCAACCTTGATGTTATCACACACATTTTTGCCTGTTTGGGAGGGTCGCCTTT
AATTGTCTTGCTGGAGCATGTATACCATTTTCCTGTCATATTCATATATAAGC
CTTGGGTCAGCAGAGTAACAGTGCAAAGATTTACCTGTCTTGTGGCTGCCTAA
GACCACACTTCTATCTGTAAGTTCCCCCAATAAAACACTCTTTGCCAACAAAC
TGGATTTGTCTGTCTTGTTCTTTGGTTTCTCAGCTCCTTTGGCACTTGAGGGCC
AATTTGTATATATGGCCCTTTCACAGAACATCAGCATCTCATGAAAATATTGC
TTCCCATCACATACAAACTTCTCTTCCAAAGACATTCTGGTAAATGTGGAATA
TTGGGTCTCTTTAGAATTCCAGTGATTTTAGACATTTTTTATTGAATTATATAA
TATTCACATATCCTAAGAGTACAAATAGATGAATTTTCACGAACTGAGTTCAT
CTTAAACCACCATCCAGATGAAGGCACATTACCGTTATGGGATCTTTGGGAT
GTCGCTTTTCTGGCTGGAAACCTCTGTGGCCGTTGGTGCCTTTGCCTGAGTTCT
TGTCCTGCATCCAGGAAGAATGAAGTATGCAGACAAGTGGAGGGTGAACAA
GATAAAGAGGAGCTTTATTGAGTGTGAGAATAGCTCAGAGGAGACCTGCAGT

GGGTAGCACCTCTCTGTAGGCAGGCTGTCCCACCGAGTAATCGGCTCTCAGC
AGAAAGGAGGCCCTGGAGAGGGTGGCCTCTCTCTGCCAGCTGGTCATCCTGT
CACCTCTGCAGCTCTTAGCAGAGAGGGTAGCTCCTCGGTGCATCTGGTCACCT
CATATCCCGCTATCAGCAGAGACAGTAGCTCATCTCTACAGAATGGTCATGC
CATCATCTCTCTATTCTCTGCCCTGCTCTGGCTGAGCCTGGGGTTTTTATGGAC
CTCAGAGGGGAGGAAGTGCACAGCAACTGGTCCATGGGCAACCATGGATGG
GCCAGAAAAGGCACCACAGGTCCCCACTCTGGTACGTGGGTCTGGAAGCCCT
GCCCTCAGCCTTTGGGACCTCCCTGGTCTGAAGGTGGAGCCTTATGGAGGAC
CCATCTCCTTCTGCCCAGGAATCTGTCTGCCTCCTGCTGCTGTTCATGGCTGCC
TAGACTCAGCCCCAACGTTGTTCCAAGATTGGAGCCAGCACCAACAGCAGGG
AAAAACTAGGCAGCAGGACAGGTTCTTTGGAGCCTGCAAGGGCAGGGGGCC
TTCCTAGGCCCACAATAGTGCAGGGATGCCTGAGGCTGCAGCCTGGCCCAGG
AGGGTGGGGCTCCCACCTGCTCCGTGGAGTTGGAGGCTTGGGTCTGCAGCGG
TGGTTTGGGTGGCTGCAGCGGTACTCCAGGAGCTCCTGCCCTAATTTGCAAGG
GGTGGGGCTCTTGCTTGTCCCCGGCTCCTGCCGACTCCATGGAACATGCAGCC
CCAGGCTGCCTGCCATCTGCAGCCAGTGTGATGACAGCAGTAAGCCATCTGG
AGTGGCCACTGCCATCATTACTGACACCACAGAAGCCCTGCTTTATGCTTCTT
TTCTGTTCTTATCCACATCCCTCTCCCCCATTCCAGGACAACCACTGTCCTGAC
TTCTACCAACATTGATTAGTGTTTCCTACTTTTATATTTTATGTGAAGGGAACA
ATACCACATTCTCTGTTTTGTATTTGGCTTTTTTGCTTAACATTATGTTTCTAA
GTTTCATCCATATTGTTGTAGTTCTAGATTTGTTCACATTTCTGCATAGAATTC
CATCGTGTGAATATATGACAATTTATTTATCCATTCTACCGTCTATGGGCACT
TGGGAAATTTCCACTTGGGGACTATTATTTAAAGTGCTGCTACAAACATGTTA
GTGCTTGTCTTTCTGTGAACACATGTATGCATTGGCATACACATATGAGAATT
TCTGGGTCTTAGGGATGGCATATGTTCAGCTTTAGTAGATGCTGCCAAACTGT
TTTCCCAAGAGGTTGTATTAACTTACAATGCTACCGATAGCATATGTGGATTC
TGGTTGCTGCACATCTTTGTCAAGACGTGGCATTTTACATCTTTTTTATTTTAA
CCATTCTGGTGGGTAAGTACTAATATCATCCCATTATGGTTTTAATTTGCAAT
TCCCTGATGACTGATGAAGAGGAACACTTTTTAATAAACATTTTAGCTATTTG
AATATCCTCTTGTTTTTTTTGTAAATCTTTTGGTAATTGTTCTGTTGAGTGTCTTT
TTTCTTATCGATTTGTAGAAATTCTTTATAAATTCTGGTTAGGAGTGACTCATC
TAATATAATTATTGCAAATATTGTCTGCAACTTTATGGTCGCTTTTTCATCTTG
ATGGCTTTTGATGAACAGAGTTCTCAATTTTAACATAATCTATACTTTTTTCTA
CCTTTTATATTTAGTGTTTTATTGCGTCCTGTCTAAGAAAATGTGGCCACCTCC
GAATTTAATTCCAGTGATGTCTTATCATATCTAAGTGATATTTTAAGGACCAC
ACGTTCAGGCGACATCAATCATAGACAGTAAGACAGGTTGATAGTTCAGTTA
CCTTTGAAAGAGGCTTAATTTTTAAATTTCCATATAGAACCCAATAGCCAACA
AATATGTCTTGGGATGTCACATTTCAGTAAAATATTGCTTCCACCTTACTTATT
TTTTTGAGTCACAGCTTAACCGTCAAGTTTAGAGGAAAACTCAAAATGTTCTT
TACTCAACTTTTCCTTTTCTTTCAGAGACAGCTTCGCCTAAAGAAAAAAGCAC
CCATTTAGTACAATGGGCTTGTGTCTGATGCTATTTAATAGCAAATCACTTTC
TGTCTCCTAACCATAGTAACCACGTCTGCAAAAGTTGAAGAATAATTCCTGCT
ATGTCAATATTGCAGTGTTGTTATGAAAATAACAACAATAACAATAGGATGA
AAAAGTGCCTGCAATGCAGTGATACTTTATGAAGTAGCGCACTGCCAAGAAT
GTTGGTGAATAAGAGTGACTGTAACACAGAGTCTGAGGCTTATAGCCTATCG
TGGTGACAGGAAGGTACCATAACCAAAATTTTCAAGGAGAAAGTTAATCTAC

CAGAGATTAACTCTTTTGATTATCTCTGGAGAAGTGTTCCTTCTTGTACCATTC
TTCCAAAGCACAGCCAAAGGCTTTATAAGTTTATATGCAAATAATAAAATCA
CACAACCAAATCTGTAAAAGATTCAGCAGGTGAATGTCAATCTTTAATATGA
TACTAACATTTATACTGTACACAAACCTATGGCTCTGTTTTGTTAGTTCCTGCT
CAGAATCTGACTACCTTTTTCACTGAATATTTTGGAAAGCTAACTACTCTTTC
AAATCCTGCCTTTTGACCAGGTACAGTGGCTCATGCCTGTTATCTCACCTACT
CTGGAGGCTGAGGTGGGAGGATCACTTGAGCCCAGGAGACTGCAGTGAGCTA
GGATTGCACTACTGCCCTCCAGCCTGGGCAACAGAGTGAGAGTCTGTCTCAA
AATAAAAATAAAAATAATAAAATTCTGCCACTGATTAAACCCATTTTCAAAA
ATTCTTTAAGCATTTCTGTGAGAGACAGTTTACAAGACCCATGAGAAAACCT
GTCTGTTTACTTTATAGTGTTATTTTTAACCAAAAGTGGCATTATCCTTCTTGA
CATCAGACTTCACTTTGAAAGACTTTAGACTGTGTCTAAAATCACAGCCACCA
TCCAAGCAAGGTTGGCATCAGTCAAGTTGTTTACATACAAGTGCACAGACAT
AGGGTTTCTGGAGTTAAAAAAAAAAAACCAACCAACCAACCAAATCATCCCCC
TCTTCCACCAAATTCCAGGACACCCAGTTAAATTTGAATTTCAGATAAACAGT
GAATAATTTTTCATTATAATTTATGTCCCATGCTTATAACAAATTCAAATTAA
CTGGGCACTTGTATTTTGTCTGGCAGCCCGACCCAGGGCTGAAAATAATTCTA
GAAGAAGAGTTTCTGAAACTGCTTTGTTAATGGCAGCATTATGAGACTAGAA
TATAGTCTCTCAAGGCAACAGCCCTCATTTACACATCTAAATTATGAGATCCT
TTTTTAAGAAAGGGGATCACCTGATCTTTCTGTAAGACTTCATTACATTGTAG
TCAAGAAAAGGACACATTAGCAGGTAGCAAGCAAAAAAGTATGTGAATTTC
ATTAGTGTTCATTGTTTGTTACACCTTGACCAGGCTCTTAAATTAGCAAATAA
GCAGTTCCTTATAACCTTTCCAAAATCTACCTATGTTTATTTAAGTTGAGTCA
GATCAACTGGTTTTACTCAAATATTGTAAGGAATAATGAATAAAACAAATAG
AAAAGTTATGCTACCACAACAACAAACAAAAAGGGAAATTCCCCATTGAAG
ATTGGTCTGTGAGGACCACTTCCTGGTCTTAACTTTGCTTCCTCTGACTCCATT
GGTAGAGAGGTACGCAAATTTCTAAGGGAGCACCCTAGTGCCTCATAACTCT
TTGGTTTAACCATCATCTAGTAATAGCCACCTGTCATTAAAAAACCCAAGCAG
TGAAACACTGCCAACACATGGAAGGCGTATAGAACTGAGAGGGCTGAGGCT
GTCAGGCATGGGAACAGGTATTTTTGTATCCTTTCAAATATTTCAGTAGTGCT
TATTATATAGTCAGTGTCTTCCTGACACAGTATGCCTTCTATTCTCAACATGA
ATTTCCTTAGAGTCACTTTTCTTTGTGCTTTCGATAGTTTCCAGTTTTATTTTTA
TTTATTTATTTATATATATATATTTATTTATTTTCTATTTATTTTATTATTATTA
TACTTTAAGTTTTAGGGTACATGTGCACAATGTGCAGGTTAGTTACATATGTA
TACATGTGCCATGCTGGTGTGCTGCACCCATTAACTCGTCATTTAGCATTAGG
TATATCTCCTAATGTTATCCCTCCCCCTTCCCCCCACCCCACAACAGTCCCCA
GAGTGTGCTGTTCCCCTTCCTGTTTTATTTTTAAAAGGGAGAACTCACCTTTCT
GTAAAAGAACATCAATAGAAGAATTTTGCCAGAAAGAAAGAAAGGCAATAC
AGAGTCTTTTCTAATTCTGTCAATGGAAATGTCTTTTTTAAAAAATACACAGG
CTTTCTTTGATTAGTCAATTTTTTTCCCCAAGAGTTGTATCTCACTGAGTACAG
TTTTTTTATTCAAGAGTTGTATCTCACAGTGAGAAAGAAAGAAAGAAAGAAA
GAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAAAAAAAAAACTTGAAATAA
AGAAACTTGAAAGAAGGTTGAGAATTGTTACTTATTTGCAAATCTTGAGTTTT
GGCCTGAGAGTGGAGAGTAATTGGAACATTGAAGATGAAAAAATTTCTAAGA
GATTAAAAAAAAAAAACAAGAAAAGAAAAAAGAAAAGAAAGAGAAACAACT
ATTAATTCTAAGAAAAGCACAAGGTTACATAAAAAAGCCAATTTCTTCTTAG

CACATTGTGTGACTTAGCTGTGAATAATATTTATATAATCAAAAATAAAATAT
GTGTTGACCATTTCTCTAACAAAAAATTATGACATATCTCTCTTGGGGAATTA
TGGAACAGGAAGTGTGTGTATATGTGTGTGAAGGTAGGGAAAGGAGGGGGC
TAGGTAAGAGAGCTAAATTTCCCTAGCAGGAAATTAACAGAAAATACCTAAG
CTTTAAAACATCAGGAAATAGCTATCATACATAGCTACGTAGACATATAGTG
GTCAATATTAAAAAGAAACAACTAAAATAATATAATGTACTTTTTTTTTTTTT
TTTTTTTTTTGCTGGTGAGTGGAAATCAGAGGTAGGGACACTATTGCTGTTTT
GGTTAAAAGCCTTATAGCCTTATTTGGCTGTTTTGCTGATATACACGTATTAG
CCTAATACAATTAAAATTTTAATAAAATCAGAGACCCTTCCAAGAAACTAAT
ATTCAAAAAAACAAATATCAAAAGTAAAAAAAAATAATAAAACAAGAGGAAA
CAATTGTCTGGGGCCAGGGTAGGGTACTACTCCTAGTTCCAAGCCAGTTTTTA
AATAAATGGACTTTAGGAACATAACCTGTTCATAAGGATGGACTTTCCACATT
TCAAACCCAGTAATGGTAGAATAACTGCCTTAGGAGATTCTAGATACAGCCT
TTCTTTCCGCCAACTTGCCCCACTGTTCGACTGTTTCTCATCATGCTGGGGGTG
AACAGCATCCTCTCGGATGCTTGTAAAGCAATGGCCTGAACAGAGCTAATGT
TTTTAGGTCCATGTAACTCCATTTTCCTAGACATAAAGTTGAGAGTTAAGTTT
TGAGGCCTAATGGTCCCTTTCCTAATCTGAGAATGAGTTGGAAAGCTCAGCTC
TCCTTCCTTTCTCTGGGCTGCTCCTTTCCAGGTGAAGGGGTTTGTGATGTTGTA
AGCGAGGAGATTTGTCAGATTAAGCACACAGGGTGCCATGGAAGGTAAATTA
AATTGACGATAGATTCAGCTTAATAAGACTTTGCAAACTTTCCAAATTCTTCA
GTCAAGCTTCAAATTTCCAGGGAAAAGTGAAAGCTCTTAATTTACTGGGTTTA
GGTTGGTCTGCCCAAGTTTTCCACTTGAGTAGTAAAATCCCTCCAACTTCTAG
AATTCGAAGAGCAGCCAGGGCTGAGCTGGGAGCATTTCTCCCATCTGTGATA
GGCCAAGCTTCTCAGGAATCAATTATCCTTCTTTTGGCATCAAACTCCCATTA
GCCAATGACAGTTCCAACATAACCACCTGGTCCAGGTGGCTACACTTAGGAG
AAGAAAGTAGGGGGTAGGGGGAAAGACCCCCACAGTGCCTTTGAGCAGAGC
TCTTGCAATGTGAAGAGATACTTCTTGCACCTTTAGAGGAAAGGAAATTCCG
AACATTTGCTTTCTGCTGTTGTTCCTTCTTCTCATCTTAGGGGTTCTGAGCCCA
TTTGCTTCAGTATAAAGATAGTTCTATGCTAGTTCAGTCACAGAGAGGGCTGG
ACCAGACCCTTGCCAGAATCTCCAGAGACAAGTGCCCATGCCATACAGAGCC
CAAGAGTTAAGAATACCCTTGCTACTCTGGTGCCACATTGGTATGACATCTCC
CTTAGGCATTCCTTAGTCCTTGCAGTAGAAGAAGCCCACCTTATCTTAAACCT
GACATGCAAAAACTGAGTGATAAAGTAAATTAAAGAATGAGTATCTGTTTTG
AAGTATGATCTTTACAAAAGGGTTCACCAAAGACTGCTTAGATTATTAGATTG
ATCAAAAAAATCAATTTATGATACATCTTTTTCAGATGCGTATTATATAAAACA
AAAGTGCTAAGACTGGGTATAACTGCTAGCAACAACAATACAAGTAGCTAGA
ACCAAAAGCATGCACCACCATGCCCAGCTGATTTTTTAATTTTATTTTAGAGA
TGGGGTCTCACTATGTTGCCCAGGCTGGTCTCAAACTTCTGGACTCAAGTGAT
CCTCCTACCTCAAGCATCCCAAAGTGCTGGGATTACAGCTATTTTATATATAT
ATATATATATATATATATATATATATATATATATATATATACACACATATATACA
CACATATATATATACACATACACATATATATGTATATATATACATATATATAT
ATACACACACATAATTTACTTTTCTCATGAGCCTTGTCTGCGTGATTAGAATA
TAAGATCCATGAAGTCAAGACTGGTGTGACACATTTCTTTTTATCTTAGTCCC
TGATATAGTTTGGATGTTATCCCACCCAAATCTCATATTGAAATGTAATCCCC
AATGTTGAAGGTGGGGCCTGTTGGGAGGTGATTGGATGAATGATGGGGGCCG
ATCTCTCAGGAATGATTTAGTACTTTCCCCTTTGTACTGTTCTCGTGATAGTGA

GTAAGTGCTCACAAGATCTGGTTGTTTAAAAGTGAGTGGCATCTCTCTCCAAC
TCTCTTGGTCCTGATTTCCTCATGTGATATGCCTGCTTCCATTTTGCGTTCCGC
CATGATTCTAACTTTCCTGAAGCCTTCCCAGAAGCTATGCTTCCTGTATTTGCC
TTCTGCCATGATTGTAAGTTTCCTGAAGCTTTCCCAGAAGCTGATGCAGAAGC
TATGCTTTCTGTACAGCCTGCAGAAGCATGAGCCCATTAAACCTCTTTTCTTC
TTAATTACTCAGTCTCAATTATTTATATTAGCAATGCAAGAACAGACTAATAC
AGTCCTTAGCCCAGATCTTACACATAATAGGTATTCAATGTACACTGTGTGTT
GTTGATATTCACAAATTACTTCCTCTCTTTCTACAAAATGTTGAATAATACCTC
CATCAGAAAAACCTGGGTTAAAGTAAGGGTATTTTTGTCTATCTGCAAAAAG
ATAAACATATTCTAATTTTTCTATAATGATTGCGGGTAGAGATAATCTCTGCC
CTCAACAACACTCTTCCCTCATAGAGGGAAATGAAACTACAAATGTATTTGA
ATATAATAATAGTGAAGGAAATAATGTATGCTGTGGTCCGTTTCCAAGACAA
AGTGCCTTGAATCGGTTTAGGTCAGCAAACCACAGAAGAAATAGGATATACT
AGGCCCCTGCTTGGATAGCCAATGCCTGCTTGTCACCACTTCCCCTTAGTTGC
CCTCACCCAAACCAAAGAAGTTTAGTCTGAAATGAAAGCTTACTAGCCTGCA
AAATAGCTCGTTTTTTCTGTTCTTATTAGCCTACCCAGCTACTTAGGTCATAAG
TCAAATACTTGAGTTCCTAAGCTAACTAGGATTGCAATGTATTGTGGGCTGCA
ACAAAATGCAGCAGGACAACCCTAAAGAAAACACCTAAAGCCACTACCCAA
CAACCGATAGGCAATGTCCAGGAAGACTGTGACCCCATAGTACTCAGCCTGT
GAGGAACCGGGGGAAGGGACCTGTGCATTAGGGAATAAATTGCTTTTTGTAA
CTGTGCTGGGTGTGCCTGCCCACCGGACAGCCAATCTTGCAAGACCATCACG
AAAAATCTCACTTTTACTGTTCTCTGGGTCTCTGAGTCCATTCTTTGGGCTTGG
ATGGTGAGTTTGTTTCTCACAATAGCAACAAATGCAACAACAACAGAAGCTA
ATTTTTATTGGGCACTTACTATATGCCAGGATCTGCTTAAAGCACTTTACATG
TGTTAGCTTATTCAATCCTAAAAATAATTCTTCTAATCACATGCCTCCACATT
GTCTTAAGAAACTCATCCTGTGTTCAAAAGCTGGATAATTTTCCAATTTTACA
GAATCAGGTTGACATACTCTACAATCCTAGTCAGCATGATAAAGTGACACTC
ATACATTCATTTGAAAGACTTTAGGGAAATAGTTACTACATTGGCACAGAGA
TGTGGTGCCTCAACTCTGTCATGAAATTAGGACTTGTATGTGTTACAAGAAGA
GGTGTGGATGAACTAAAGAAATTGTTTCTATGAGTAGAGATTTTGAAACAGA
GAGTCCACTGGATCCCAAGTCACTGCTGTGTGAACTCACTCACCACACGGGA
ATTCTCCAAGTACCATCCTGCCTGACTCATTAATCTTATGAAGCACAGAGTGA
TCACACATGCCCCTGAAATGACTGTATGTAAAGTAAATTCAGGCTGTAACAC
ACAAAGTTTTCAAGGTTGGCCTCATAGCACTGTAGATTCCCCCAGCAGATGG
GCAGAGGAAGGAAACTTACTCTGTCTGAGATTCTCTCGTATTTCCAGGGCAA
CAAATCATGCGTAATGAAAACAAAGCAAAGTCAGTACCAGCCCAGGGCCAG
CCATCACCCCACCCAAACCAGAAGGGCAGGAGCCTAATTCATGAAATGTGCT
GTGCTTCTTCCTCCGGCCAGCCAGGCTCGGGGTTTCCTGATGTGTTCCTGGAA
CCAGAGCTAATGGAATCAGGAAAGCATGTTACTTTGCCACTGCCAGTCATTG
CAAGTACAACAAAATAAATATTGCTTTAAAGAAAACAATTATCATAAAGACA
ATTAGTAATGAAAACAGTTATGCCTTCCTTTGTTTTTGAGACAGGGTCTCACT
CTGTCACCCAGGCTGGAGTACAGTGGCACAATCTCTGCTCACTGCAACCTCCG
CCTCCCCATTCAAGCAATTTTCGTGTCTCGACCTCCCGAGTAGTTGGGACTAC
AGGTATGCATCACCACATCTGGCTAATTTTTGTATTTTTTGGTAGAGACGGGG
TTTCAACCATGTTGGTCAGGGTGGTCTCGAACTGCTGATCTCAAGGGATCAGC
CCACTTCAGCTTTTCAAAGTGCTGGGATTACCAGCGTGAGCCACCGTGCCCGA

CCCAGTTCCGCCTTTCTAAATTGGCCTCTTAATATTTTAGAACATTTCATTCCT
CTGGCCTTGAGTGAAGAATAGAAACTACAGAGGGAAGGATTTGGAGTGGCTA
ATGTTGGCAGAAGTGAGAATCAGAATTATGGAACTGCAAAGTCCTATGACCT
TCCATTTACTGAAGAGGAAACAGAAGCACAGCAAGAGTGCTCAAGAGACTTA
CCTAATGCCACTCCACACAGTAAGTACTGGAATCCGGGACTTGGACTGCCAA
TTCCATGTGCTTTCATTTGTGACATTACTTTTTTTTTTTAAAAAAAAAGAACA
ATGTAATGTTTCAATAAAATTTAAAATTTTGGTTAAAAATCACCTATAATCAT
AACCTTCTGATAACTATTATCATTCTTGTATGTTTCCTTCCAACATTATTTTCA
TACATTTAACATAATCATAGCCATTAACCATATGTGTTTCTGTTTTTTTTGTTT
GTTTGTTTGTTTGTTTTTGAGACAGAGTCTCGCTCTGTTGCCCAGGCTGGAGT
GCGGTGGCGTGATTTTGGCTTACTGCAACCTCGGACTCTCAGGTTCAACGGGT
TCAAGCAATTCTCCTGTCTCAGCCTCCCGAGTAGTCGGGACTACAGGCGCCTA
CCACCATGCCTGGCTAATTTTTGTATTTTTAGCAGAGACGAGATTTCACCATA
TTGCTCAGACCGGTCTCAAACTCCTGACCAAAACGAAGTGTTTTTTTAACCTA
CAATTTAAGTAAATGTTTAATGTTTATTTATAATTTATTTATTAAATGTTTACA
TTCACCAGGCACTATTCTAAATAAAACAATGTTGCTGTCCTCACATAATGAAG
ATTCTGGAGGGGTCAACATAGAATACGTACATAACCAAGTAATGTACAACAT
TCCACCATGATGAAGATAATTTGTTCCTAAATAAAGGTCCTCGTGGGTGAATT
CTCATAGGCTGAAAATGTACCTAACATTCATTTTGACAGAAAAGACACTCTTT
TCTCTTAAGCCCCCAAATTAACACCTATTTATGTTAAAATAACATCAATCCCA
TTAAAACAGGTACAATTATCTCAAAGGTAAATGGTTATCATAGGACTGATGT
CTGTGCTCATAAAGCATCAAAGCAACCATATGATTCTCTGCGTTTATGTAATT
AAATGTTTAATGAAAAAACAAAAACAAAAACATGAGCTCTTTTTTGTGGCAC
CTTGGAGGCAATTAGCTGCTTCAGGATGAAGCTAAATATCTCCTCCCCAGCCA
CTGGCTGACAGACACTCATTGGATTGGACAACGAATGGCAATTTTGTACTTAT
GAGAAGCATATGGCACAGAGGTTGCTGCCGACGCTCTGGAAGAGTTATGTGG
TCCGAGTCAGTGGTGGGACCAACGAACAAGGTTTCCTCATGAAGCAGTGTGT
CCTGACCCATGGCTGGGTCTGCCTACTGCTACTGAGTAGGGGCATTCCTGTTA
TAGACCAAGGAGAACTGGAGAAAGAAAGTGCGAATCTGTTTGGGGTTTCATT
GTGCATGCCAATCTGAGCATTCTCCAGTTGGTTATCTTAAAAAAAAAAAAAAA
AAAAGGAGGGAAGGATACTCCTGGACTGACTGATACTATGGTGTCATGTTGC
TGGGGGCCCAAAAGAGCTAACAGAAATGCAAACTTTTCAGTCTCTCTAAAGA
AGACAATGCCCCCCACCCATATGTTGCCAGAAAGCCCTTAAACAAAGAGAGT
AAGATACTGAGGACCAAAGCACCCAAGATTCAGCATCTTGTTACTCTACATG
TCCTGCAATACAAACACTGGTATTGCTCAGAAGAAACAGCATACTAAGAAAA
ATAAGAAAGAGCCACAAAATATGTTAAACTTTTGGCCAAGAGAATGAAGGA
GACTAAAGAAAAACTCCAGGAGCAGGTTACCAGAGACACAGGTTGTCCCCTC
TGTAAGCTTCTACTTCTGAGTCTAGTCAAAAATAAGATTTTTTGACTTACTTA
ACAAATGAGTAAGATCATACCACCCCCAGCAAAATAATCAATAAAGACTAAG
GACATTGATGAAAATGATGGAGAAGATGATGACTTCCATGAGTACTTTAAAG
GAGAAACTTAAAGCAACTTGTATAATTTTTAAGACTGCAGAGCTGACATGTG
GTAACTTATGCTGAATCTGCTCAGTAGCTTCCCACAGCCCCCAAATCTGAATC
ACATTCCTTCAATGTGAAGAATGCCTCATAAGCTAAAGTTGAGGAACTAAAT
AAACTGATTAAGTCTCTGTCTTATGACACCTCCCACCCCCTATCACAGTTCTA
TCCAGGGCAGCATTGACTAGCGTAATAGCAGGGTGTGATATGTGGTAGGACA
GATAGGATGGGGAGGACAGGGAATGATGTTAGGTGGAGATGGGGGAATGAC

AAGGAGTCTGGTGTGAGGGAGCCCACTGAGGGAGGGCTAGGTGGTAGAAGC
TAATGCTGGCCAGGTAGCTAGGTGCTAGATTAGAGAGCACCTTGCTAGCCAC
ACTGAAGACCTGAGTACTTTGGAAGCCAGTGGAAACTTCTGAGCAAGGAGTG
ACACAATCAGACTCATATGTGAGAAGGAGACTGTGGGTGAGGGAGATAAGTT
AGGTGACTATTGTGGGGTTCAACTGAAAGAGGAGGTAAAAGCATGGGTCAG
GGTTACAATGGTGCAGACAGAGAGAAATAGTTATAGTTTATCTGTAGTTTGA
ATATAAAACCAACCGGATTCCCTTTGAATTGAATATGATACATCAGGGAAAA
AGAGGAATCAATCTTAGTGCATACTCTATGTCAGGCCTTGTTCCTATGTTTTA
CATATATTCATAACAATAAAAATAATATCAAATGCCTTTCTAATACCTTCCCT
ATGCCAGATTCTATATGCTTTAATCTACTTTCACCCACATCACCCTTACTGCA
ATCCCATAGGGTAGGCACTATTATCATTTCCATTTTCTGATGAGAAACTGAGG
CACAGACAGGAAGGGCAATGCACACGGGAAATGCTGGAGCTGATTTTAGAC
CCCAGACGTTCTGGCTGCAGAGTTGAGATTCACAATCACTGCAAACACTGCA
TTCCATGGTCACAAGGTTTTGGTCCGAGCAGTTGAGTGGGTGATGTTGCTATC
GACCAAGACAGGAATACCTGAGGCAGAGCAGGTTTGGGGGAGTGGATAGCG
AAGGCCTGGGTTTTTGGCCATGCTAAGTTTGGGATGCCTGTTAGACCTCTAAG
TGGAAATGTTGTGTGGCCATAGGGTGTGAAGAATCTGGAGTTTAGGAAAGGA
GCCAGAACTGGAGATAGCTGAGCTATATGGAATGATCAGCTCATAAACAGAA
CTTAAAGCTGCGGGACAGGATGAAAGTACTGAGGAAGACTAAGTCCTGGTCA
CTCCAGCATTTACAGTCCAGAAATAGGAGGAGCTCCAGCAAAAGAGATTGGG
AAGGAGTGACCTGTAAGGCTGGAGGAAACCAGGAGAGTGTGCTGGAAGGTA
AAAAATTTCTAGGAGGGAATGATCCACTCTGTGAAATGCTGCTGAGAAGTCC
AGCAAAGGGAGGATGCTGCCAGATGTCATGATGCTCAGTTTAATTGTTGCAT
ACAACTTCTTCAAGTGGAAGAATTCTCTTTTTATCTACTCTTGCATTTTCAACC
CCTTCACATATAGCTCACAAAGTAGAGGAAGAGAGCTCATCTAACTTCAACG
TGAAGTTGTTAATTTGAATTCAGTTTAAATATTTATTGGGTGTCAGGTATGGT
ACTAGGCCACAGAGATTCAGGGGTAAGTGAATCACAGAGTTCTTGCTTTTTTT
TTTTTTTTAGACAAGGTCTCGCTCTGTCACCCAGGCTGGAATACAGTGGCAAT
CACGGTTCACGGCAACTTCTATCTCCTGGGGCTCAAGCAATCTTCCCACCTAA
GCCTACCAAGTAGCTGGGATTATAGGTACATGCCACCACATGGGTATTTATTT
TATTTTTTGTAGAGATGGGGTCTTGCCATGTGGCCCAGGCTGGTCTCTAACTC
CTGGCCTCAAGTGATTTTCCCACCTTGGCCTCCCAATGTGCTGGGATTATAGG
CATGAGTCACTATGCTGGGAGGGTTCTTGCTCTAGACAATGGACTAAGCAAT
GAGAAGGAGGAGGGGAGGGAAAATAGGAGGAGGAGAATGATGAGGAGGGG
TGCTGGGGAGGAGGGGAAAGGAGAGGAGAGGGAGAAGGAAAGGAAAGGGA
GAAGGGAGGAGGAAGAAGTGGAGAAAGAAGAATACCATGATAATATTAATA
TTAACATTTGGAGGTTGCCTAATGAACTCAGCCAGAACACCCCATAGAAAAG
CCCAATGGTTGACACACCTCCGCTTAACAGGCCATAAAGTACTAGGGGGAGA
ATAAAGGCTTTTTCACATATGCTCATTCTCAAAATAATGATCTCCTGTGCATG
CTTTCTCAGGACGCTACTGGATGATATGCACGGCCAAGACAAGGGAGTATCC
AAAGAAGAAACATGAGATTCAGGCAATAAAGAATCAGTAGGATACAGGCCA
GAATAATTCCCACCATATTTTAAACTGAGATCCCAAGACAAACACTGTGCAC
CAGGTCTAGATAACAATCAGTCCACATTGATTGCCTGATTTTCAAAACATATT
AAGTGGAGGTTTACGCTCTTGGGGAAGAGTTTGGGAAGGAGATTATGGGGGA
AGGTTGCAGGGGGTGTTTCATTTTTGATTTTGTTTCTGAGTATGTAGACATTCC
CTTAGTTCCCCAGTGTTCAATACAGAGGCCGCCTCACAGTTACATCAGTGTTC

CCTGTCCGGAGTCTCCTGTTAAGTTCTGTGAGTAAGACTTCAGTCTTCTGCTA
GGCTAGGGGAGGGACAGTTGCCCAGCTGCACAGGGTAGGCAAGGGGATAGG
TGAAATGTCCCTTCATACAGATTTGCAACCAACTCTGTTTTTTAAGTCTCTATC
CCAACATAGGCCCTGCTTCAAGAGCTATCACAACATCCAATTCCTGCTCCTTT
TGGGGCTTCTGAGATACAAATAAGCCTGCTTCTAAGACCTTCTCATGCTGCTG
TCTTGAATTTCAGCTTTCTCTGCTTTTCTAAGTGAGTCTCTTCTTATTCACCTG
CTTTTTGGATCCCAGAATTATATGAGTGTTTTGTTTTTTGTCTCTTACTCTCTTT
ATCCTTATGACTAGAGTCATTTTTGTGAAGTTTGGAAGGGAACAAAAGGTAA
AATATGTGCTCAATCTGCCACACTGTTTAAAAAGTGGTTTTCTTTTTAAATTA
CCAAATATATGCATACAGTTTAATATATATTATAAAGTGAATGATCATTAACC
AACACTCAAATCAAGAAATAAACATTGCTAGCTTCCCTAAAGCCCCCCATAT
GTCCCCACCTGATTACAACTAGTTCTCTCCCTTTAGATGTGACACCATCCTAC
CTTTTATGATTACTGCTTCCTTCTTTTCCTTTACAGTTTTACCACCTATGTATCC
ATTGTGAAACAACTGAGTTTAGCTTTGCCTGCTTTTGAATTTTATGTAAATAG
AGACATATGGTGCATATTCTTTTGTGTTTGGTTTCTTTTACTCAGCTCTGTAAA
ATTCAACCACATCATCATTACATATATCCACTTTGGTTATACACGGATACCAC
AGACTATCCATGTTACTGTGGATGTACTTCTGAGTTGTTTCTAGTTTGGAGTA
AATCTTAATGCTATGAATATTCTTGTACACACTCTTTGTGCACATATACACAC
GTTTCATTTGGTATGCCACAAGAAGTGGAAATTCTGTGGCACAGTGCATAAA
CATCTTCATCCTTCTAGATATTATCCGTTGTTTTCCACAGTGCTTGTACCAGCA
GTCTATGCAAGTTCCCAGTGCTCCACGTTTGTGCCAACATTGGTATTGTCTGA
CTTCAGTGGTGGCTGTACAACCTAACGTGTTCAGCTGGAAATACCTTGTTAAG
TTTTCCATGTACCTTCTATCGCTTCATCACTGCACTTTTGGAAAGAGCTTCTAA
AACTCCTTTCTTATAGAGGGCAGACCCATAAGGCAACCTACATGTTCCATTGT
TTTTTCTTAGAGAAGATTTCAGGTAGAGCTTCTGACAACCTGCTCCAATTGGG
ACTAGCTGCTCACTAATTCTGCCTCAATCATCCGTTTCCTGAATCTCATGTCTT
TTTTTTTTCTGCATTTTTCCCGTTTTCATAGAGCACATCTTCCAGAAGCTTCA
TGAGAAATGATGTGTAAGGAAATACTATTTTTGAGACCTTGCATATTTGAAA
ATATCTTTATTCTCCACTCATAGTAAATAGTTTGGGTAGAGAATTTTAGGTTG
GAAATAATTTTTCATAAAAAACTTTGAAGGCATTTTTATTTTCCTCTAGCTTTC
GATGTTGCTGTTGAGAAGTCTGATGCCATTCTGATTTCTGATCTTTGTATATG
ACCTATTTTTTTTTCCCTCTCTATGGAAGCTTTTAGGATTGTCTCTGTGTTTCTG
ACATTTCACAACAGAAGTGAATTGTTAAATTCACTACACACTTGGTGAGCACT
TTCAGTATGAAGAATATGTTCTTAAGATTTTTGGAAATTTTCCTGAATTTTCCT
TTGATAATTGCCTTCCCTCCACTCTTTCTCTAGAATTCCTAGAGTCATGTTTTC
AACCTCCAGCATTAAACCTCTAATTTTCTTTTCTTTCCACTTTTATTTTTCATCT
CTGGGTATTTTTGTTCCACTTTATAGAAGCTTTCCTCAATTTTATCTCCTAACC
CTTCCACTAATTCTGCTCCTATATTTTTTAGTTTCCAAGAGCTCCTTCTTATTCT
TTATATGTGACATTTTAATAGGAAACTCTTCACATTTAATGAAAGCAAAATTG
TCCTCCTTTTTTAGAAAATGGTAACTAGTTTTAAAATGTTTTCTTTTGCCTCTG
CATTTACTGTTTTCTCAATGTTTCTTTTACAGTTTATTTTGGACTCTGTCTCCTA
TGTGAAAGGTTTTTTCCAAGTGGCTAGTGATCCCTAACTGACCAGTTCATGTG
AAGGGGAGAGGCACCATGAGCTAATCAGAAACTTATGCGGAGGGGGGACAT
ACAGATTAGAGGGTCTCACTTTAGGATAACCAGGTGGGAGCCCTGTCTTTTTA
TAGCCTCTAAATGCCAGTTTTTGTTTTGTTTTGTTTTTTGTCTATTCTTTTGGGC
TGCTTGGTTTTTCTAGAAAGGAATTCTCTCCAATCCTGCCTAAAATAGTGTAA

GTCTGGTATTTAGCATTCTGGGAGCCTGGTAGGGGAAAGGGTGGATGGGTCA
GGGATGGTGGAGTGTTGTCTCACTCTTCATTAGGCGAACTTCTCTGATTTTTA
TTTTCCAGTAAAATGCCTTGCTCTAGCCCTTAGCTGTGACTAGTGCCCTTAAG
CCAGAGTTTTTTGGTTTTTGTTTAGTTTAACCTCTCCAGAAAGGGTATCTTTAG
TTTTCTGTCAAGAAGAAGGGAGGAGCAGTTACCTGGCTGCCCAGTCTAGGAG
AGGGGAAAGGATGTGGTCTCTAAGAACTCCGTATATGAGTCCTTGGGTTTTTC
ATTCTACCTCATATCTCTGCCTTTAGAGGTATACAGCATATCTGATTTTGGAT
ATTTCTTAGGGGATGTATTAGTCCGTTTTCACGCTGCTGATAAAGACATACCC
TAGACTGGGTAATTCATACAGAAAAAGAGACTTAATGGGCTCACAGTTCCAC
GTGGCTGGGGAGGCCTCACAATCATGGTGGAAGGTGAAAGGCAGGTCTTACA
TGGCCCCAGGGAGAATGAGAGAGAAAGAGAGAGAGAGAGAAAGAGAGAGA
GAGAATCAAGCAAAAGGGGTTTCCCCTTATAAAACCATCAGATTTCTTGAGA
CTTATTCACTAACATGAGAACAGTATGGGGGAACTGTCCCTATGATTCAATTA
TCTCTCACCAGGTCCCTCCCACAACATGTGGGAATTATCAGAGCTACAATTCA
AGATGAGATTTGGGTGGGGACTCAGCCAAACCATATCAGGGAGCACTATAGC
ACAGATTGTTTTGTTCTTGTTGCCTTTCATCTTTTATGGTATTTAACAAATGAA
GAAGGCTAAAATTGAAGTTAATTTCCACTTGTTTGTCTGCATTCATTCTTCCA
AAATTGTATTGAATACAGTCGATAAATTGTATTTATCTGCAGTCACCCCTTGT
CTCCTTCTCTTTGTCCCTATAAGCTAAACACCTTTTTTATTCCTTTAATGCCAT
TTTAGTGGAGTATATACAATCCATACATTTTCCTAGGTATTTTTCTTTCTTCAT
TCATACTTTCTATATCCAAAAGAGGATTTGAGCTTGTTGCAATAAAATATACA
TATGCGAAAAAGTTGAAATTTGGGAAAAGTAAAAAAATATCAAGTAGTAAAA
GAAAGGAAACACCTGTGGTGGAAATCTAGGCTAAGGATATAGGCCGTGACTG
TACAAAGGTTGGCCCTTACTAAACACCTTGGCAGTTCTGCTAAAAGGAGGAA
CAGGAGGAATTTCTCAGCCCTCATTATCTAACAGGAAGCACTCCAGGGCATC
AGATAAAACTTCTGCTAAGTTTACTGAGTGAGTTGATTCTGTAACTGAATAAA
AGTTCTTGGTGCTCCAACTTGGAATTGATTCTACTGAAGAGATTAGAGTGAGA
CCCAGATAAGAAAAATAAATAAATAAAGATAGATGTGAGGGGTAGGGATGC
TAAGTCTTGATTGATTGGCATCTTTCCCAACTCAAGCACTGTGATGACACCAT
CTGTTGTTACTCACTCTATATTCCCAATTTATTCAGGGTTTCTAGGTGGAAGG
ACATTAAGAAAGTACCCTGTAGAGGAATTCTTATAGTCTTCTCTTTCTTTTCC
ATTTCAACAGGATATTCTCAGAGCCCTCTAGCAACTGTGACTGTGATTTCAAG
GCAAGAAGTAAACAGACAGTAGGGAATTATGTTAGGAGTATATATTCCTTTA
CTTTTCCTCTCAAAGAGAGAGAAAATAGGCTTTTTTTTTTTCCTGATAAGAAA
GTGTATGAGTAAGCCTAGATTCAGGGCCCAATAAGATCATGCTCAGATTTTC
AAATTACAGTTTTAGAAATTTTGGGGGAAATTCTTGATGCCAAAAGGCATTGT
GAAATAGTATCAATACAATTGCAGGTTTAGCATTTTCTTTATGCAATAAACAG
TACAGGGCTCTGGGGCTTTACAAAGATGAGCAAGCAACAGTCCCTTTTCCCA
AGGAGCTTACAGTGTCTCACAGGAGATAATAAAATGACACAAGTGATTAATA
GAGGAGAGAGCTTCAGAATGTCTAGCCAGCTACACATGTCTAGCCACCTGCC
TACATGGAGACCCAGTCCCCTTTTCCCACCCACAAGGTGAATCTGGGAAGCC
ACAAACAGGAACCCTGCCCTTCTAACTGCAGCTGATTAAATTAGAGGTCAAG
AGATACGTGACCTAAGGAAAACCAATTGGATTTCCTCTCCAAAGATTTAAAA
TTAGAATTCAAAGGTGCTAATCAGTCTCTGCTGTCCACTGATTTAAGGGCGTA
GAAGCACTGGCTTGGATATTTCTGACCAGGCACCAGTGGCAATGCAGAGAAA
ACACATCTGGAGAGAAAGAGGCATGCAGAGTGGCTTTCCAGTCCTGCTTTTG

GCCTCATGGGATCCATGAGATACTCCCATATTAAAAGTTTTGATTCAAACAAC
TGTAATACAATACTACTGTGAAGACTATGAAATATATGCTAAGGGGGCTCAA
AGAGGGAGAGAAAGCACATATTATTGTGGAGCAGGTGGGATTAGAAATAGT
GAAATATTACATGAAAAATTTTTACGACTGAGTTCACCTTAAATGATGGGTA
AGATTCTAACAGGAGGAGATAGGTTGAGAGCATTAATTAGTGAAAGAACCAT
ATGGGGGAGAGAGAAACCAGGTGTGTATGTTCCAAGGTGCCTGGGGGTGGTT
AGGTTTTATGATGAAAATTAAATAGCTCTTGGAGTAGATTCTATCTTTGCATC
AGCTATGACGCAAGTCAGCATATTCTAGGTCCTTTCTCCACTTGGAAAGAATT
ACTGCCAATTATCAGTCCATTTCCATTGGCTTCCCTCTAACTACTTTGCTTCAA
AAAAAATGAAAAGTTCATTTATTCCTATAGTTCTATAAAAGAAATCTACTCAA
AAAGATGTGAAATGACTTATAATGCAATACTGTAATTTTTTATATAAACTTCA
TCTTTGTTTCCTTGTTACAGGATAAGTGGTAAGTAAATATTGCCTAGTAATGT
GACATGAGTAACAGAAAATACAAACTTATTTTCGCCCTAGGGAAGCCTGCTT
ACTTTTCTTGACCCTCTTCTCAATATCTATCTATAATCTTCAGATATAGCAAAG
GGCCAGCAACCACTTTTCTGGGAAAAAAAGATATTTTGCCAAACTTTGAAAA
CACACCAAAATATGGGACTGAAAAATAGTGCATATATATCAATTGAGTGCAG
TGGTCTGATCATAGTTCACTGCAGCCTCGAACTCTTGGCCTTAAGGGATCCTC
CTGCCTCAGCCTCACAAAATGCTGGGATTACTGGTGCCTTGCCCTATTGTTTA
AACTAACATTTTTCATAAAATACTAAATGTGAATATCTTCCAAAACTTGAAAG
AACTATGCAGTTATAAAGCATTATAAAAATAGGCATATTAGATATTTTTATAT
GTTTTTAAGTTCATTGATTAGCGAGAGGAATAAAACTGAACTCAGTAGAAAA
GTTTTGGAGAGAAACAAAAAAGTGAGGATTTTTACCTTATAGCTAACATTAT
CTACCTCATTTAGAGAAGGATCTTGTTTTTATACTATAATCCTTTTAGACAAC
AAGCCAATGAAATTTTAAATTCAAAGGCAACTCAAATGATTCTTGACAAGGG
TGACAAGACTTTTCAATGGAAAAGGGTAGTATTTTAAGCAAATAGTACTAGG
AAAACTGAATATCTACATGCAGAAGAATCAAGTTGGACCCTTACCTAACACT
GTATACAAAAATTAACTCAAAATGGACCAAAGACTTTTTAAGACCTAAAATG
ATACAATTCTTAGAAGAAAACATAGGTCAAGTCTTGAAGATATTAGAGTTGG
CAATGATTTCTTGGATATGACACCAAAGGAACAGGGCACAAAAGTCAATAAA
TTGGATTGCATAATGATTTAAAAATTTTGTGCATCAAAAGACACTATCAACAG
AGGAAAATGATAACCCACAAAATGGGAGAAAATATTCACCAACCATATACCT
GATAAGCGATTAATATCCAGAATATGTAGACAAATCTTACAATTCAACAAAA
AACAATTTAAAATGGGCAAAATACTTAATAAACACTTCTCCAAAGAAGATAT
GCAAATAGCAATAAGCACATGAAAAGGTGCCCAACATCACTAATTATTAGTG
AAATGCAAATCAAAACTACAAGATACCACCTCACACCCATAGGATGGCTACT
ATTTTTTTTAAAAAGAAAATAACAAGTGCTGACAAGGGTGTGGAGAAATTAG
AATGCTTGTGCACTGTTGGTGGGAATGTCAAATGTTACAGCCACTGTGGAAA
ACAGTATGGCAGTTCTTGAAAAAAAAAATAGAATTAGTGTATGATCCAGCAA
TTCCACTTTTAGGTATATGCTCAAAAGAATAGAAAGTAAGGATTTATGAAAC
ATTTGTATATTCATGTTCTAGCAGCATTATTCTCAATAGCAAAAACATGGAAG
TAACCGAAGTGTCCACTGACAGATGAATGGATAAGCAAAATGTGGTATATCC
ATACAATGGAATATAACTCAGTCTTAAAAAGGAAGGAGATTCTGACCTATGC
TACAATGTGGATGAATCTTGAGAGTATTATGCTAAGTTAAATAAACTAGTCA
CAAGAAGACAAATGCTGTATGATTACACTTATATGAGGTATTTTGAGTATTCA
AAACCACAGAGACAAAGTAGAATGGTGGTTGTAGGGGGTTAGGGGATGGGT
CAATGGAGAGTTAGTGTTTAATGGATATAGAATTTAAGTTTTACAAGATGAA

GAGTTATGGAGTTGGACGTTGGTGATGGTTGCATGACATTATGAATGTGTTTA
ATGCTACTGAATTGTACACTTTAAAATGGTTAAGATAGTAGATTTCATCTTAT
GTGTATTTACCACAATGAAAAATATTGGGAAGGCCAGGTGCGGTGGCTCACG
CCTATAATCCCAGCACTTTGGGAGGCTGAGGTGTGTGGATCACCTGAGGTCA
GGAGTTCGAGACCAGCCTGGCCAAAATGGTGAAACCCCATCTCTACTAAAAA
TACAAAAATTAGCCAGGCATGGTGGCAGGCACCTGTAGTCCCAGCTATTTGG
GAGCCTGAGGCAGGAGAATGGCTTGAACCCGAGAGGCGAAGGCTGTAGTGA
CCCAGGATCGCACCATTGCACTCCAGCCTGGGTGACAGAGAGATACTCCATC
TCAAAAAAAAAGGGGGGGGGGAAAGTAAGTCATATTTCAAAATATACTCAA
GAAATATTTTCTTCAAAGGATCTGAGCAGCCAAGTAAGGGGCAGTCCACTGT
GTCAAGATGCCATACCGTTCTGGTGCACTGTGTGTCTCTCCATTCTGGGTCTG
CCTCTGCCTTCCACTTCACCGTGCTTCTGCCTCAGTGGTCCTGGTGCCAGTGG
TCTTCTAAGTAGCTCCCTAGAAGAGAACAGCTGTTTGCCACAACAAGGGAAA
AACTGGAGACTATCAACTTGAAGTAGAGAAGGTAACAGTAGGAACTGGTTAT
TGTCATAGTGTTTTTAAATCTGAGAACTTTTAACTTTGCCAACTAAGGATGAT
TGATGGGGAAAGAGTAAAATATCTGGGAAACTGGTCTCTGAGAAAGAAATTT
GTTCGGTGGAGGTTGTGGGTGAGATGGATGCGGCAGAGACTGCTACTATCCC
CTAGGTTCCATTCTCTCCTTCTTCTTTAGAAATAGGACCTTCAGTTTTTAGTTC
AGCATGTGGTCCTCTGGTACAAAGGCTACATTTCCCAGTCTCCCTTGCAACTA
CATGTGGTCATGTGACTAAGTTTTGTCCAATGGGATGTAAGTGGAACTGCCTT
GTAAAATTGATGGTAAAAGTCCTTAAAGGGCATAGCTATGCCCCATCAGCCC
TTTCCTTTTTCATACTACAGTGGAAGTCCAATGTGAGGATAGCAGAACTGTTA
AGAGAGGAGTTTTAATCCCAAACAACTTTCTATGGCCTCATCCCCTGACCTAC
TCCAATCCTGGGCTACCTATTTCTGTACTTTAAATTGAAAGAGAATACATTTG
CATCTTGTTTAAGCTATTGTTATTTTGGTTTCTGTCACTCAGCCAAACTCAATA
CTACTTGATAAAATTGGTAAAAAAAAAATGATAAATTAGAAAACTGCTTTAT
GTCTAATGAAGCTTTAAAATAAAAAGTACTTCCTCATATGGGTTCTCTGTTTT
TCCTCCAAAATGTCTATGAAGACACAGAAAAAGAGGAAATTTACAACATTAG
AAATTAGGACCAGTGCAACTTAGTTGAGAGAATTTTCTAATAATTCCTTCTTT
TGTCTATTCATTTGTTCAGCATGTTTATTAAAAAGTACCCCTCATGTGCTAGG
CTGTTCTAGGCCCAGGGAATATAGTGGTGAACAAAACAGATAAAGTTCTTGT
CTTCATGGACCTTCTATTCTAGTGGGAGAACAGAGACCACCATCAAGATAAA
AAAAATAAATATAATGTCAGTTTGTATATGTAGATGAAGAAAAACAAAGCAA
AGGAAAGGAGTAGAAAATGATAGACACAAATTGTGAATTGAGTGGTCAGAG
AGGGCCTCTTGGAGGAGGGTGACATCTGAGCAGATCCCTGAATGAAATGACA
GTGGGAGTTCTGGTGATATCTGGAAGAAGAGCATTCAAAGCAGCGAGAACA
ACATGTGCAAAGGCCCTGAGACAGAAACAGGCTTGGCAGATTCCAGAAATG
GTAAGGAAGACGATGTGCTTAGAGAAGAGTAAGTGGATGAAAAGAAGTGGT
AAGAAGAGATGTCAGAGCTTGTCAGGGGACAGACAGTGTAGGATAAATTGA
CAGGAGATAAGTTGATTAAAAATCATACTTTGTGTCCACTCTGAAAAGAAAA
TGTTGTGAAATGATGGGAAGACCTCTTGTCCCTCCATTACAATCTACAATGGG
TTCAGAATCATAACACCTACTCTCATGAGAGAGATGGAGTATTAGTCCATTTT
CATGCTGCTGATAAAGACATACTTGAGACTGGGTAATTTATAAAGAAAAAGA
GGTTTAATGGACTCACAGTTCCATGTGGCTAGGGAGGCCTCTCAATCATGGTG
GAAGGTAAAAGGCATGTCTTACATGGTGGCAGACAAGAGAGAATGAGAGCC
AAGTGAAAGGGGTTTCCCCTTATAAAACCATCAGATCTCATGAGACTTTTTCA

CCACCACAAGAACTGTATGGGGGAAACCACCCCCTTGATTCAATTATCTCCTA
CTGGGTCCCTCCCATAACACATGGGAATTATGGGAGCTATAATTCAAAATGA
GATTTGGGAGAGGACACAGCCAAACCATATCATTCTACTCCTGGCCCCTCCC
AAATCTCATGTCCTCACATTTCAAAACCAGTCATGCCTTCCCGACAGTCCCTC
AAAGTCTTATTTCAGCATTAACTCAAAAGTCCACTGTCTAAAGTGTCATCTGA
GACAAGGCAAGTCCCTTCCACCTATGAGCCTGTAAAATCAAGAAGCAAGTCA
GTTACTTCCTAGATACAATGGGGGTACTGGCATTGGATAAATATACCCATTCC
AAATGGGAGAAATTGACCAAATAAAGGAGCTAAAGGCCCCATGCAAGTCCA
AAATCCAGTGGGGCTGTCAAATCTTAAAACTCCCAAATGATCTTTTTTGACTC
CATGTTTCACATGCAGGTCACACTGGTGCAAGAGGTGGGTTCCCATGGTCTA
AGGCAGCTCCACCTCTGTGGCTTTGCAGGGTACAGCCTCTCTCTTGGCTGCTT
TCACAGGCTGGCATTGTCTGTGGCTTTCTCAGGCACATGGCACAAGTTGTTGG
TGGATCTACAATTCTGGGGTCTGCAGGATGGTGGCCCTTTTCTCACAGTTCCA
CTGCGCAGTGCCCCCGTGGGGACTCTGTGTGGTGGCGTCAACCCCACATTTCC
CTTCTGCACTGCCCTAGCAGGTGTTCTCCATGAGGGCCCTGCCCCTGCAGCAA
ACTTCTGCCTGGACATCCAGATGTTTTCATGCATCTCTGAAATCTAGGCAGAG
GTTCCCAAACCTCAATTCTTGACTTCTGTGCACAAGCAGGCACAACACCACAT
GGTAGCTGCCAAAGCTTGGGGCTTGCACCCTCTGAAGCCATAGCCCAAGCTG
TACCTTGGCCCCTTCTAGCCATGGCTGGAGCAGCTAGAACACAAGACACCAA
GTCCCTAGGCTAGACACAGCAGGGGGTCCTGGGCCTGACCCACAAACCATTT
TTCCTCCTAGGCCTCTGGGCCTGTGATGGGAAAGGCTGCTGCAAAGTTCTCTG
ACATGGCCTGGAGACATTTTCCCCATTGTCTTGGAGATTAACATTTGGTTCCT
CATTACTTATGCAAATTTCTGCAGCAGGCTTGAGTTTCTCCCCAGAAAATGGG
TTTTTCCTTTCTATTGCATCATCAGGCTGTAAATTTTCCAAACTTTCATGCTCT
GCTTCCTTTTTAAAACTGAATGCTTTCAACGTCACCCAAGTCACCTCTTGAAT
GCTTTGCTGCTTAGATATTTCTTCTACCAGATACCCTAGATCATCTTCCTCAAG
TTCAAAGTTCCACAAATCTCCAGGGCAGGGGCAAAATGCCATCAGTCTCTTT
GCTAAAACATAGGAAGAGTCACCTTCACTTCAGTTCCCAACAAGTTTCTCATC
TCCATCTGAGACCACCTCAGCCTGGATTTCATTGTCCATATCATTATCAGCAT
TTTTGTCAAAGCCGTTCAACAAGTCTTTAGGAAGTTCCAAATTTCCCCACATT
TTCCTGTCTTCCGAGTCCTCCAAACTGTTCCAACCTCTGTCTGTTGCCAAGTTC
CAAAGTCTCTTCCACATTTTCAGGTATTTTTACAGCAGCACTACACTGTACCA
GTACAAACTTACTGTATTAGTCCATTTTCACACTGCTAATAAAGACAAACTCG
AGACTGGGTAATTTATAAAGAAAAAGAGGTTTAATGGACTCACAGTTCCACA
CGGCTGGGGAGACCTCACAATCAGAAGGAGAACGGTATGGGGGAAACCACC
TCAATGATTCAATTATCTCCCACTAGGTACCTCCCACAACATGTGGGAATTAT
GGGAGCTACAATTCAAAATGAGATTTGGGTGGGGACAGAGCCCAACCATATC
AGTTGGGCAGCCAGCCATCTGTTGTTGGGTGATTTTTACCAAAGTGCATCTTT
CTCTGTACAATCAGAGGCTTCCACCTCCAGAGAATGTGTTATCTTAATCCTCT
GGCTGGATGCTTATGAGGATAGAGTAAACTGGCAATGCGCAGTGAATCTTGG
GAGGGGTCATCCTTACAAATAAAGCACTACAGTTTGGTAGAGATGGAAGAAC
TTCAGATGCAGTGGATGCTGTAACCGATGGCTTTTATGATGGTATTTCATGGT
GATAGTGTGAGAACGCCAGACAGAGGGAACAGGTGACAATATTGTCTCTGTC
CAAAAGAGTCATAGTAAATGCAAATCTGCTTATCATGCTATTGAGTTAGGTA
CTAAACTTTGCTATGCCATCCAGTTGTAAAACCTAGAGCCCCAAAGAAGAGA
ATAATAAGATAGAAAACAACACAACTGATCTTTTGGCTAACACTAAGTCTGG

AAATTACCCTTCCATTTGAGATGATTGCCAGAAAAATCATCAATATATATCCT
ATGAAAAACTCCTGAGGGGCAGAGGAATGGGATCAATGCTTCAAAGACTCA
GAGAAAGAAAATCACTATAAAAATTATTTGCTCCTAGAATGTAAAAAATATT
TTAAGAAAACTGCTTGGCATTCTAAAGGTGAAAAGACTTTAGAATGAGATAA
AAGGGAGAAGATTACTTTTAAAAATTCAGCGACATTAAAAATGTAATGACAG
AAGTAAAATCTACATCGGGAGTAGGAATGTAGACTCCTGTAGCATTACTATG
ACAGACACCAGAGTCAATGACATGAAAGACAAACTTAAGATCTTTTCAGGTA
TGTGGTAAAAGAAAAAAGGATTCAAGTGAATGATATAGAGGACTGGGGCTG
GGCATGGTGGCTCACGTCTGTAATCCCAGCACTTCGGGAGGCCAAGATGGGC
GGATCACGAGGTCAGGAGATAGGGACCATCCTGGCCAACATGGTGAAACCCT
GTCTCTACTGAAAATACAAAAATCAGCCGGGCGTGGTGGTGTGCACCTGTAG
TCTCAGCTACTCAGGAGGCTGAGGCAGGAGGATCGCTTGAACCTGGGAGGTG
GAGGTTGTAGTGAGCCAAGCTTACACCGCTGCACTCCAGCCTGGGCAACAGA
GTGAGACTACATCTCAAAAAAAAAAAAAAAAAAAAGAAGAAGAATATAGAGG
ACTGGATAGATTTATTTAAGAATTATACATTTCTGAGGAAGGTACTAGAAGA
ATTAGAACAACAATAATTAAAGATATCACTTTAAAAAACAAGACCTGACTAT
TCAAATTGAAAGGACTCACCATTCTAGATGATAGTAATGAAAAGAGGGCTAT
TCTAGAAATACCCTGGCAAAGGTTTGGGAATGCAAGTATAAAGAAAAAAATC
ACATCTATTAGTATTCTGTCATAAAGCAAGTAATTTATGAGGAAACAAAAAT
CTAGCTGTTCTCAGATAGCTCTCCAGTAAGAAATGCCAGAAGACAACAGAAA
AAATACTTAAAATTCTTTGAAGAAAAAGACTATGACCTAAGAATTTTCTCAG
CTGAATTTTTTTTCTCATATGTGAAGACAACAAATGAACAATCCCATATATTC
AAAGGCTCAGGAAATAGAGCATTCATGTATTCTTCATGAAAATATTATTTGG
AGACATACATCAGACACCCGAAAGATGTGAAAGAAGAATAGGGTATGGATT
ACACATAGTTCATGGCATTATGTAAATGTTATAAAGCAAGATGACATAGCTT
GGGAAAACAAAAGCTGTGTCTAATAGCAGTACTTCCAAGCCATAACTTACAG
TAGCCCAATTCTCAATAAATTGGAAGGCAGCTAAACAATCATACAGTGCTAG
TATTTTATAGTGTCAGGGTCTATTCACATATAATCTCATTTAACCCATTTTAAT
TCCGTGAGATAAAAACCAATATCCCCATCTAAGATATGGGAAACTAAGACGT
AGAAGAAAGCACTTGGCTAAGATATCATGGCTCGTAGGTGGCAGTCAAGAGG
TCAGTTTGCAGTCTACAGATTTAACCTCAGACTATTCTGCTTCTAACATGACT
ATAGAAATGTATTGATCATTAGCTGCCTGAAGTTCTGTTCTTAACTCTAGGTG
TCCAAAAGAAGATGAATTTTGTTTAGATAGCATTTTCCCATATCTAGGGCTGT
TGGCTTTAAAGAAGTTCCCAGAAGTGAATCCAATTCCCAGAAGGATCAAGGT
GGATTCTTTTGTGTGTTACTCTAACAGGTTGCTTTATATATATATATAAGTTAT
ATAATATACACATATAATTTATGTATTATATATAAGTATGTATAACATCATAT
GTAATATATAAATATAGTACATTTCACACTGGTAGGACGTAATTCCAAACCA
CATATTTGTTTAACCAATTGAATTACTTGGATTTAGTTTCTATTTCCCTCTTTA
TTTCCTTCTGAAAATATTATACAAATAACATCAGTTTACAAAAAATAAAATCT
GATATAGCTTTTTTATCATACACTAGGCTAGACTAAATGCATTCTGTGGATTG
TTTACCTAGGACCAGATGGTATATTATAAAGTTGTATACAAATGAACAAGGA
CTGCCTGAAATGGATTGATAGCTGAGCACATTTGGCTGGAGCGTCTATTTTAG
AAGGAAAACCTGAGAATCATATTATTTGAGTGTGATTCATGTGTTAATAGTAT
TTCAAGACAAGCCACTTAAAATATGTCCTGAGTGGTGATGCTGGAAATGATC
TTTTCTTCAGTGTTTTCAAGTGTCTTCATTTAAGTGTACACATTTTGCCTACCT
ATAACACGTATCTACACATTGTGATTAAGAGAGCAAACTCTGAAATCAGACC

TAGATTAAAGATCTTAATCTCTGCCAGTTGATTGGGCAAGTGACTTAATCATT
CGACAGTAATTTCTGCATTTGAAAATGCCTATCTCAAGGGTAATGAGAATTAC
ACTGGTTCACCAAAAGAATTACACTGACTAATATAAAACGTGCCAAGCATGT
ATGTGTCACATGAAGGCTCATTAAAAAGTGGATATTATTGTTAATCTTCCAAT
AACTACTATTTCCAACAACAGGCTGAAGGGGCTCAGAAACGTTTGTTGAGTA
AAAACACAAGGAAACAGTAGCACAGATTTCCTGCTCTCCTTTACGATCGATG
ACCTGTCTAAGGACTGTGATCTCTGTTCGCTACAGATTGTCACCTGCATTAAT
CTACTGTCACCCATTAACCTATCAAATAAGGCAGTCTAAAAACTCCAGGCGT
CCCTTTCCGTAAGGACCCGGACTGTTGAACTGGAAAGCTAAAATTCAAGGCG
TGACAATTGCCCTTTGTCCCACATTCCTCCACCGGTCGCCTGCTTATTTAAATG
GTGCGTCCCCTCGGGTACGACTTGAACAAAACCTGCCCAGAGCGCTCCCTGT
GTAGATTCGGTGGAAGCAGCTGGAGGCTCCAGTTCTCATCTGCTGAGGTGTCC
CCGGCCGCCTTGGCGAAGTCGGCCACTCCAGTTCCTCACGTGGTGAGCAGTCA
GGGCAGCGGGTCGATTTCCGAGGTGCGATACCTGGGTTTGAGGGGCGCGGC
TCGCAGCGGCGGGTCCAGGGGCGAGTGCCAGCCCTCACCCGGCCTCGGCGTG
CGTTCGGAGGCCGAGACCTGGAGGACGCCTCGAGTCCCGGCGGGACGCCACG
CCTGCGCGCCAGGCATCCGGGATAAGAAGTGCGCGCCGGGCTCCGGCTGCGC
GCCGCGGGGCCACCAGTTTGCGCGCAGGGCTGAGGCGACCGTGCGGCGATGG
ACACGCCACGGGGCATGGGCACGTTCGTGGTGTGGGACTACGTGCTGTTCGC
GGCCATGCTGCTCATCTCGGCGGCCATCGGCATCTACTACGCCTTCGGTGGGG
GCGGCCAGGAGACCTCCAAGGACTTCCTGATGGGCGGCCGCAGAATGACCGC
AGTGCCCCTGCGCCTGTCCCTCACCGCTAGCTTCATGTCAGCGGTCAGTGTCC
TGGCGACCCCGTCGAGGTCTACCGTTTTGGGGCCATTTTAGCATGTTTGCC
TTCACCTACTTCTTTGTGGTGGTCATCAGCGCGGAGGTCTTCCTCCCGGTGTTC
TACAAACTGGGAATTACCAGCACCTACGAGGTAAGGGGCAGGGTGGGCTGG
GACCATGCAGGGCGCGGGGGAAGGGGACTCTGCAGACCTCTGGAGGCGTTTT
CTTGGGGGCAGACTGTCACTGCCACATCGAAATCTCTCCCCGTCCATCGTCAT
CATCACCCTCCCTTCACTCCTCCCCGTCCCCCGATCTTCACCTGGCATCTTTCC
CTTCTACTGAGAGGCGTCCTCTAAGGGTGAAAAAATTCTTGGGATTTACTCTC
CTGGGCTTAGTGAAAAAAAGAGGCTTCCAAAGTGAACGGATTGCAACAGTAG
TGCTCGCTATGGTCCTTTCTACCTTTAGCATCTTTGATTCCCAGGCAAGGGGA
AAGATTTTTGGGGAAGGTAAGTTCTTCAGGTCTCAGGCCCTGCTTCTTGAAAG
GAATACTCTTGTTCCAGGTCCTCAGCCCCATTCACTCTTCCAAGATAGTTGGT
AAAGAATTTGTACTCCCCTTCCTTCCCCTTCCACACCCACCCCCTTTCCACTGA
AACAGGCACCCAAGTGGCTAAGGTGTACCAGTACCTGTATTTCGGACCAGAT
TCTACCACTTACTAGCTGGGTGACCTTGGGTGACTTACCCAACCTTTTGTGCC
TATTCCTCAAAAATAAAATAATATCACCTACCTCATCCAGTTGGGAGGATTAA
ATGAGATGATGCAAGGTGTAGCATTTAGAATAGAGTTTGACACAGAGTAAGT
GCCAGATAGGTATTGGCCATAATTACTGTGGTGGTGCTAGTTGTGATGGTGGT
AGTTATCTTGGGATCAGTAGGAAAATTAGGCAGCAAGGTTCTCAAACAGCTC
TGCCTTCTCTATGGAAAGACAATTAAGGAAAACTGTCTTTCCTCTATTGCACA
GGTCCCCTGGCTATAGGTTTTCTCAAGTTGTCTGCAAGAAAGGGCTTTGAAAA
GATGATGCACGTTTAGTTTAAAGAGCTGTTGAATGGGAGGTGGGGGCGGGAT
CAAAGCTCAGGGGAGTGGCGGGGCCCAGTTCAGTGCCTTCAGTGGTCTCCCT
CACCCAGCACTGGTCCTTGAGCTACATGGGACAAGCCTGCAGAAGTTTCCTCT
CTATTAGGCCATTACCCAGAGGACTCCAATCCTGAGATTCTGGCAGGTGGTTT

CCTAGTGAGGCTGGTTCAACTGGTGCTGGGGACAGGGCCTTAACGGCACAAC
AAACATTGGTGACACACTCTGCCTGGGCTCTAGTGACATCTTCCAGGGCGAG
GCACCCCTGGGCAGCATTGCCTTCCATTCACTCTGGGGCTGTGTTTGGGAGAG
CTCAATAATTATCCCAGAAGTGAGCAGGATGCAGTTTCTCAGGTTTGCTCAGA
AGCGACAAACTGGCGAAATTCCCCAGACCAAATTCACTTCTCAGTCATTTTTT
TTTTTTACTTCTCCTTATAGTGAGAAGAAATAAAATATAATTTCAGACTATGA
CTTCTTTCGAGAACTCAAAAGGCTGGTAACACTGGGCCCATGCTCCCTGTGGC
CACAGGAACTCAAGTTTATAGCTATTGCTCTCACTCTCTATTAAACTTCTATA
TTGAAGTCTTATCTTTGTGTTTTCATGGTTACTGTTTTTTTGGTAGAGAAATTT
TTATCTGCATTTATGTCTTTATCAAAAAATGTGAAAGGCAAAATGGACAGGG
GTAAAGGGATCTCTTGTTTCCAGGGAGATGTAAGAAAACATATATCTTAGTG
GGTGGGAGGAATATTCCAACATGTTAATATGTGAACACCTGGCCACTGTCAC
TTATGTATGTTACCTGATGGAACCTGGAGGCATCTGAGTCTGTGAGCTGTGAT
CTATAGCTAACCAATTCCCAAACTTGGAAAGGGTTTGAAGAATCTGGCAGGA
GGACTTCAGTCCCCACACATGAATGCTGCTCTCATAACCACTGACCCATGCCT
CTTCGGGGCTACAACATTAACTCAAGCAGAAACAGATTTCCCTCATGCTTGCA
GTGGTAGAGGCTAAGTTAGCAGGGGTTCCAGCTAAAAATGTCAAAAAGCAA
AAGTTCATTGAATGCTGAACACATGCCAGGCACTGGGCTGCATTGCATGTTTT
ACATACATTATCTGACTCACATTCAGAACAATCCTTATGAGGCAAGTACTATC
AAGAGCCCCATTTTGTGGACAGAGACTCAGAGTAAGTTAAGCTACTAGGAAT
GTGAAGCAGCGTAGACATATGCCTAGCAGCATACTGGCTGCTTCTGAAGGCT
CTGATGAGTCTGTTATAACCCCAAGCAATGCTTACGCCAGTGCAGAGAGAAC
AAGGGAAGCAATATAGGACACGTGAAGTTGTGGGCACACTGCAGGCATCCA
CCTCCAGGGTGCCAGGCAGTTGGGAACGGGGAAGTTAGGCAACATGCTAAG
ACCCTCCCAATTGGTCATGGCAAAGACTTTCCTGTGGAGTATATTGTTTATTG
TTTTATGCTGAGAGGTACAGAATCTATACAGAATATGCTGTACAGTGTTGAGT
GATTGATACCCTGGCATTCTTGCTCAAGTATGAAATCTCTCCCTGTTGACATT
TCTTTAGTGAAAATTAGCTGTTCACAAGCATTTTTTTTCATCCATCACTGCCTT
CAAAATGTAGTCCTGATAGGTGTTTAGCAACTTCCAGACCCAAGTATAATGC
AGGAATCCAATTTAATCCACCAATTCTTCAAGATCCTTCTGGAGCCAACTAAA
TCTGTGTTTTAAATTGAGTTGCTTAGACCTTATGTTTCTTTGTTATTTATATATT
TATTCATATCAACAGAAGCAATTCAAAGTTCTGAGGACCACGAGAGATCTTG
ATTAATTTAGGCACCTATCATTACTCTGTAATTATGTCTAGAAGTCTTACAAG
TATTCCCTAAGGCTGCTGCACAGTGACTTGGGCATCTGGTCCAGTCTCCGGGT
ATTTCTTCAGCAGCCTAGTAATTGATGGCTGAGATCGTAATAGACTTCTTGGT
TCTGTCTTGAGTTGGTCCTTGGGCTTAAATCACAGGTTTCCCTGGATCAGTTG
ATTAACTTACTTGCTTCAGTTTGGTTCAATTTATTTGGGTCTCCAGCAATTTGG
ATTCATGTACCTAGAAAAAACAAAGACTTCTCTTCTATTGATGTCCATTCACA
TGTTTAGTGATTCATAGAGATTTACAAAACAAACAAACCAAAAATAAACTTC
TAAAGAAATATAATAATACTGCCTTGTTCCCAGACTTAATAAACTTCTCCAGT
CTTCTAAAAAGGCTTCCAAAATATCCCTGAATGATTTTGACACTTGTAATTTA
TTGAGATTCAATTTTCTCCAGGCCCATTGATTTGGATACCAATAGGTTCTTAA
TATCTTACTTTTTGAGTTTACATTTCCCCATCCATTTATGTCACTGAATAGTCT
CCAAACCCTGCAAAATGAGAGGGAAACAAATTATTATTATGCCCCTCATTTT
AAAGAGAAAGAACCTGACATTTGGAGAATGATTGCCACAAGCTGATGGAGTC
AGGCGTGGGACTGGGTCAAGGGCTGGAACTCTTCCTCCAGTGTACTACAAGG

CACTGTGTATTTTTGTCACCGATCAGTGATCTGGGAGTGAGGTCTTACTCTGT
GTTAGTTTCACCCATTCACTTATTGAGTCTTAAATATGTAGTCAGATTATTTAC
TTATATATGCAGAAAAATGTGCACAGATTTTGGCATCAAAGGTTTTAAATGTG
TACTGTGGCCCTGGCAAGTTTGCAAGGTTTTCAACTTGATGTTCTATCTGAGG
CCCAGTGTTCTTGTCTCTAACATGGGTTTATACACAGTACCTGGCTCATAGAG
TTGTTGGGAGGACCATGCACACAGCAAGTACTCTGTATGTGTTAGCCATGATC
ACGGGGCAAAAAAAATCAGATGTCTTTTCCTCTGATTATTACCTAATGCCAGTT
TTTCCTTCCTCTTCTTTTCTACTCTATGAAGATTGTTGTCAGCCTTAAACTTCTT
AGTAATTATATTCAAGTTAAAGAAAATCTTAGCTCCCAGCTCTGAAATTCTAT
GATCCTAACTCTGTAATATGTATATTATAATACATTATATAACATAGTAGACA
AGCAGACTATGTTATAAGACATAGCAGATTTCAAGTAGAATAGGAAAAATTG
GCATGGTCATTGGCACTGAACAGAAACTGACTTGGGTGATTTAAGTTTAAAA
AAACAATTGTTAGAAGGACATGGGGGTAGCTCATAGAACAGAGGGACAGGC
TGAGTAAGCAAGCCTTGGAAAGGACAGAAACCGGGGCAGCCTTTTCAGGCCA
CTCTGTCAGGATCAGTCATTTTTTGTCTTTGTATCCTGCCTTTAAGATTCAAAT
ACCCAATGGAAGGAAAAGCTGATTGGCTGAGCTGGGATCATGTGATTTCCTC
TTGGTCAGAAGAGGCAGGGCAATTGGCTGACTATTCCATCACAATAGTCTGC
AGTCAGAAAGGGTTGATGCCCCAGGTGAGCCAAGAAAAATGTATGGAGTAG
ATACCAAGAGAAAGAGAACAACAGATTGCTCGCTCTAAAAATACAGAAAGT
AATTTCATTTATATGACATTCTGGAAAAGGCCAAACTAGAGAGGGAAAGCAG
ATCAGTGATTGCTAGGGTTTGGCGGGGGAGAAGAGGGCTTGATCATGAAGGG
GAAGCCCGAGGGAGTTTTCTGGGTGATGGAACTGTCTTGTGTCCTGATTGTGA
CAGGGGTTACATGAATCAGTGTGTGTTAAAACCCATTGAACTGTACTCTAAA
ACAAAAAGAGTCAATTTTGCTGTGTATAAATAAAAATAACACTAAAATAAAAA
TACAGAATACAAATCAGTTATGAAGTTGCTTTACATTTCTAAATTTAAATTTT
CTTCTTGAGCTGCTGATTTTAAAAAAGGCATCCAGAGGATTCGCATAATTTTT
TTTTTTGCAACAGTATTTAGAACTTCGATTTAACAAATGTGTTCGTCTCTGTGG
AACAGTCCTCTTCATTGTTCAAACAGTAAGTAGCTCTCCATTTATTTTATCTTT
TGTCAGCTAAAATGTTTTTAACCATGGCATCTGGATTAAGCTTACCTGGGAAT
CACATAAAAGACAAAAAAGAGATGTTGAAAAAATGAGGGAACAAAAAAGG
AAAAACACTTGTCACTGTACAGCATCAACCCTTAAGATCATCAAACAGTGTTT
CAGTAAATGCTTTACTTCCTGGGATTTAAGTGAATGTTAAATATTATACTGAT
AAAGCAACAGTCGGAAAATATGCTTCCCTTTATCTGACCTCCCTTCACCTCCA
CCAAGGTGACAGGAAGGGCATTACTACATGACAAAGATATTTGTTGCTGACA
GTGGGAATTCTAACAAAAGGAAAACAAGTAGCATGTTCACAGTATTTCTGTA
ACATATTAATAGGTATGAAAAAATTAACTTCCATAGGAAGACAGTAGGAAAT
AATTTTCTTTGAAATCGTACTTTAAAAGTCAGTGCTCTTTTAAAAATTAGACT
ACAAGAACCAACTAAAGCAGTCCTTAGAGTGTAAAACAACAGAATCATAAA
CTCTGGAGGACTCTTTAAAGCCAAATACTCAATCCAGTAATTCAAGAACATG
CGCTTATAGATCTACTGATTGACACAAAGGGAAAGCAAGGATTTGCCAAGTG
GTCAGTGACAGAGAATGCTTTCCACTGTTCACCGTGCTTCTGGAAGATTGTAA
TGATCATTGTCATGACTATTTATATACACATTTTCCTCTTGTCAGTTAAGCACT
TTAGGGCTGGGTTTGCTAATGGAGGTTGTGGAAGAGATTTGCATTCTTGTCTC
TAATTGCAATTCCACTTCTCCAATCAAAAGCTACCTAAGGGCCAGCCGCGGT
GGCTCATGCCTGTAATCCCAGCACTTTGGGAGGCCAAGGCAAGTGGATCACC
TGAGGTCAGGAGTTCAAGACCAGCCTGGCCAACATGGTGAAACCCCATCTCT

ACAAAAATACAAAAATTAGCCAGGCATGAGGGCAGGTGCCTGTAATCCCAGC
TACTGGGGAGGCTAAGGTGGGAGAATCACTTGAACCCAGGAGGCAGAAGTT
GTAGTGAGCCGAGATCATGCCATTGCACCCCAGCCTGGGTGACAGGGCAAGA
CTCTGTTTCAAAAAAAAAAAAGGAAAAAAAGCTGCCTTAGGATTTGCTGCAGT
GAGACAGAGTGCCTTTGTAAATTATGTAACTTGACTCCATTTTATATCTTTGC
AAATTATATAACTTAAATTTTATCAGTCCTTAACAACTGCAGTGTAAAAGGAA
GGAATCCTTTGGTGTCTCTTAGAGACTTGAGCCTGGTAGCTTGCATTCACCAA
CTGTTCAGAACCTCATTGGATCTTTGTTAGAGATGCCAACAGAAATCAGAAG
TAGGGATAAGTGTTAGGAAGGTGGCCTGTGGTCATGTTTTTAAATCTTCAACT
TGGACAGAATAATGACTGTGGAAAGTTAGTTCATTTTTGCAAAAAGAGGGGA
GCTTTACCACCTCCCATTTGAAGGACTTCATAGCTCTACTCATGTAATATAAT
CAAACATTCAAAGGTACTGAATAGATTTTTATTTTCATAATATGCTTTTATAG
AATAATCATGGAATTTGCTTTTATGGGATATATTTGAAAGATCAAGTGCAATC
AAAATTACATTTTGAGAAAAAAGACCGTATTTATCTTACTCACTACTGTACCCC
AGATCAATAAATAGTTAAGTGTATGAATGAAAATAATGAATAATATTAAAAA
GATTGAATGTGGTTTTCATTGGCTTTCAGGATTTTTTTAGTGCAAATTTATACT
TTTGTTTAATTTATGAACAGTAAAAGTTTAGAAATAGGCTTTCCAAATTTTTA
CTATTTTCTTGATTAATTATGCAGGGATTAACAGTTTGAAGACATAATTGAGG
GGTCATCCTCTTTTATATTATTATTATTTATTTTTTTCTAATTATTTCAGATGTA
GGAAATTTTATTTCAAGCTTGAGTTGGTGGATCAGTGACCATTTGCACTAAGC
ACCATATAAAAGTCCGTATTTTTACATAAGCCGGTCACAAAAAAATATTTGTA
ACTTATGACCGGTCATACCGTAAACAGAAGAGTCAACTTTACTTAAATATTTT
GCAAGTTACAAACAAATTTTATTAGGTGTTTTGAAACTGTTGTTTTAAGTCAT
AATTGAAGTTATAGGAAAACAATCAATATTTTATAACTCAGACGTAATTCAT
GAATTTTATAATTCATATAATTCATTTGGCTTTCTTTCTCCCCCCTAGATTCTG
TATACTGGAATTGTTATTTATGCCCCTGCCCTGGCTTTGAATCAAGGTACATT
TTAGAGTTGCCAGTTAGGTAACTCACATTTTGGGGTTCACTTTCAACAAGCCT
TATTTTCTCCTTGGGGAGATGGGGAGATGGAGGAATGCTTCTAGTAACCTGC
ATCAGCTTTACTTAGCGGGGCAGGATGGGTTCAGTGTCTACACTAGCTTCTTT
GGGTTTTTGGATGGACAGCTCCAAAGTGTCTGTAGACCACAGAGGTGGACTT
CTCCAGGTGGTACTTCTCTGGGATGTGCCCATCAGCCCATTCACTCCTTTAGA
ATTAAAGCTCCCTGTAGCCAAAGTCAGGATTGACGGCATCCCCTCTTGTGAAT
CTATAACCTGGAGCCCATCTCCAGGAAGCTTCCCTGTAATTCTGCTCAGGCCT
GTTGTAAAATCTGGCAGGGAGAAAAGCTTTTCTTCTCCACACTTCTAATAAGG
CCCAAAATGAAAGAGAAAGAGAGCCATGTGATTTGAATGATCAGACTGCTCC
TAGTGACAAAAGGACAGATGTCTGTAGTGCCCTTACAAATAAATTTAGGAAG
ATTGTGCTGCTCAACAAAGTATCTATACTCTCTAGATTTGGGAAGATAAATGC
AGTGAGGCTGGGATAGTTTATTGAAGCAAACGTTCATGCTAGTCATAGTTTCA
AAAGGCTTTGGGGAAACACCATGCCCTTTGAATTCTTATCTATTCGAAGTGAA
TTTCTCTAAAACGTCCTTGTAAAATGGACATGTGGACTCTGTGATGGGAAGA
ATGTGGTACAATTCCTGGGGTTAAAATGACATGAAGAAAACCTACTAATTCC
ACACTCTGTTTTCTTGATTTTATGATAGACATGACAGTAGTTACCACCTGTTTC
TGAAGTGAACAATATTATTACCAAGAGGAACTTCATGTGTAAGGTGCTCTTG
AACTCTGAATTCTGGGCATGTTCCACATCGGTATTACCAACATCACAAGTGGA
TCATCTCATTTGTCGAGAACTGAGTTATAAACTTCATGAGTTACCTATTTAGG
ACTTAAGTGTAATTGAACATATTATGGTTTTAAAATGCAGTTCTGGGAATTAC

CAGAGGACTGACTTTAATCTGTGAAGAAATCAAGCATTCTGTTCTTATCAGGC
TCAAACCACTCCCTGAGAGTAAATTAGAATGAAAGTGAAGGTGTTTTGATGA
CTAGAGTGTCAGTTGTGTGTTTTACTGTGACCAAAGCTGTAGGTACAGGCATA
GACACAGAACTTAGTGTGCAATAGATGCTCAAAAAAGCTTATTTAACTGAAT
TGAAAATTAACATTCTCTAGGTCTAACCTCTTTTTTTCTTTTTAAAATTTTGGA
AATCCTCATACAAATACTAAGAGAGAATAGAGCCCCTTAATGTGCTCATTCA
TCACTCTATGTTAATATTTATCAACTAATGGTCCATTTTAATTCACTCCCCCTT
ACCTCTAATAGATTATTTTGCTGCAAATATGTTATGTCATTTCATCTGTATTTC
AGTATGTCCTTTAGAAATAAAGACTTTAAAAACCATAATCATACCATCATTGT
ACCTAAAATAGTAATATTAATTTCTTAATTTCATATCAAGTCAGTGTTTACAT
TTCCCTGATTTTTTTTCAGTGTTTATAAGAATCAGGATCCAAATAAGCTTATGT
GATTGCAATCAGTTGATGTCTCTTAAGTTTCCCTTTTCTGCTTTTTAAAATTGA
AAACCAGTTTTTCTTTCTTTCTTTTTTCCTGACCGCCTGATGCCTGTTGAGAAA
ACTAGTTTTTATTGAGGTATAGTTAACATACAATAAAATGCACAGAGCATGG
GGAAGTGTACAGTTGGATGAGTTTTAGTAGTTGCCTAATAGCTTGTGTGACTA
CTACCCCGCTCAAGATATAGACTATTTCCGTCATCCACTCTCAGAAATTTTCC
TTGCATGTCTTTCTAACAAATCTCCTATGGGCCCAAGAAATCATTTTCTGAGT
TATGCTACCATAAACTAGTTTTCCCTGTTGTTAGATTTTATATAAGTGGAATC
ATATAGAGCCTTTTCATGTCTGGTTTCTTTTGCTCAGCAAAATGTTTTGAGATT
CATTCATGCTGCTACATATATATCAGTAGTTCATTCCCTTTTTTTTTTTTTTTT
TTTTTTTTTTTTTTTTTTTTTTTTTTTTTGCTGGTAGTACTTCATTTTATGCCTCTA
CCAACATTTTTTTATCCATTCTCTTGTTATAGACTTGGGTTGTTTCCAGTTTTA
GGCTATAGTGAAGTAAGGGTGCCAAGAACATTCTTCTAAAGGTTTTGTGCATT
TTGTTTTTCTTTTTCACACCTGTTTTCACTTCCATTGGAAATGAAATACATGGA
TGTAGTATAAAATTAAAAATATAATTCTATTTGGACCCAGCTGAGGATAGAA
ATGTGTGCAAATTTACAAGAAATTGCCAAACAGTTTTTCCAAGTGGCTATATC
GTTTTCCATGTCCACTACGCAATATGTGAGAATTCTAGTCATTCCACATCCTT
GCCAGCATTTGATATTATTCTTTTTTATTTCAGTCATTCTAGTGGGTATTACCT
ATTATTTTATTGTTGTTTCAGTTTTCACTTTCTATATGACTATTATGGCTTAGCT
TATTATTAGACTATTATGGCTTAGCTTATTGGCTTCTGAAATAATAAAGTGAA
TAAAGACAAATAACCTAATAAAAAGTGAGCCAAAGACTTGAACACTTCACAA
AAGCAGATATTCTTCTAATACATTACATGATATATCTATTCATTTACCTAGGC
CTTTAAAAATTTCTCTTAGTGCTCACTTTGGCAGCATGTATACTAAAACTGGA
ACAATACAGAGAATATTAGCATGGCCCTTGCACAAAAATGATATGCAAATTT
GTGAAGCATTCCATATTTTTTAAAAAAGGAAGAAAAGAAATAATAAAATAAA
AAGTCTCTTAGTAATGTTTTGTAGCTTTGGATGTTGAGGTCTTTTACTATTTCT
TAGATTTATTCCTAGATAATTGTTGTTTTTTATGCTATTGCAAATGTCATTGTT
TTATTTTTCAATTATTCGTTGCAAGTGTATAAAAATACAACTGACTTTTATATG
TTGATATTCCATCCTGAGAACTTGCTGAATTTGCTTTTTAAATCAAGCAGATT
TTTGGTGATCCCTTAGGATTTTCATGTTGTTTATGAATAGTGACAATTTTACTT
CTTTTCTAATCCTTATAGCTCTTAGTTCTCTTTCCTGTCTTATTGTAATGTTAAT
GTGGAAAGTTACATTGACTGAATTTTGAATATTAAACCAATTTTTGCATTCTT
GGGATAAACCCTGCTTGCTCATAATGTATTATCCTTTTTATATATTACCGAATT
CGATTTGCTAACATTTTGATAAGGCTTCATAAATGAATTGAAAAGTGTTCCCT
CCTCCATTTTCTGAAATAGTTTATATAAAATTGTGTATTTATTTCTTACATGCT
TGGTAGAAATCACCCTGAAGGCATCTGGGCCTACTCTTTGTTTTTGTAGAAAG

ATTTTGGTTTATGGATTCAGTTTATTTCGTAAATATAGAGTTATTCATATTTTA
AAAGTCTTGTTGGGTCAGTTTTGGTAAATTGTATTTTTTAGGGAATGCTTCCA
TTTGGTCTAACTTATCAAATTCATTAGCATAAAACTGCATATAATACCCTCTT
GTTATGCTTTTAATGTTTATAGTATCTATAGTGACATCCCTTCTAATTATTCCT
AATATCGATAAGTTTTCTATCTTCTTATTTTGATATATCTTTCCAGTATTCTAT
TGATGTTTTAAATCTCTTCTAAGAATCAGTGTTTTGGCTCATTGATTTTCTTTG
TTTTCTATTTCATTAATTTGTGCTATTTATTGTTTCTTTTCTTCTAATAACTTTG
GGTTTTCTAGCTTCTTAAGGTAGGCGCTTAGGTCATTGATTTTAAATATTTCTT
ATTTCATAAAGCAAACATTTAAAGTTATTAATTTTTTTGAAACACTGCTTTAT
CTGCAGCCCATATATTTTGATACCTTGTGTTTTTATTTAGTTCTCAAAATATTT
TCTAACTTCTCCTGTGATTTCTTTTGCCCATGGGTTATTTATAAGTGTATTGCT
TAATTTTCAATATTTGAGGATTTTCTATCATGTCTTTTTGTTATTTATTCCTAA
ATTAATACTGCTTTGGTCAGAGAACTAATATGATTTTAATCATTTGAAATTTA
TGGATTTGTTTTATGGCCCAGGATCAGGTCTGTCTTGAACATTCCATATGCTC
TCAAAAAGAATGTATATTGTGTGCTTTTTGGATGTAATGTTCTATAAGGTCAA
AGGTTTAGTTGGCTAATAGTGTTATCACATCTTTGATATCTTTATGGAATTTTT
TTTACTTGTTCTACCAATTGCTGAGAGAAGGATGCTAAAATCTCCAGATATAA
TTATCGATTTTTCTATTGCTGCTATTTCTCTGTCAATTTTTACTGAATGTATTGT
GAGGCTCTTTTAAAGTCTCTTTTTAAATTTACAATCTCCCTTCTATTCCTTTTC
ATTTAAATTGCAACATATTTGCAGAAGAAACTGGGTTATTTGTTCTGTCCAGT
TTTCCAGATTCTAGGTTTAACAATTACATCTGTATGGGGTCATTTAATGTTCA
CCATTCCTCTGTATTTCTTGTAAACTGGTAGTAGTTTACTAGTAGTAGTTAGAT
CTAGACTCTTTTTCAGACTCCACTTCAACTTTTTGGAAATCATATTTCATAGGC
AGTATCTATAAGTTTATTTCCTTCAGGAGTCCCAGAATGTCATTTGAAAAATT
ATCAGCAGCTATTGATGATCATTCTAGATTCATTATTTCATTGAGGCTTATAA
AATGGAGTATTCTAGTTCTGTGAATTCTTCTTCATTAGTTAGATGGAATACTG
CTATAAAGAGAAACTTCCTCTTATCAACTATTTGCTTTCTCTGAAAAATAGTT
TGTATAGGAAAGGGAGGATAAATGTCAGATTTTTCCCCTTATGAATTTTAAAA
ATAATAAATTGGTTTCCTACTATCCCCCAAGGGTGACTGAGATTTTTAAAAAA
TTCTTACGAACTCATGGATTTAAAGGTTTGCTGTTATTTCATAGCATTATCGTT
AGTATTATTATTGATGCTCAAATTGTCCCATTTTTATCCAGCCGGAGCCCCCTT
AAGTTCCTCCTGATTTTAGCCTTGGCTGCATATTGGACTCACTTGCGGGGTGG
GGGAGCTTTCAAAAATACCAATACCTCAGTCCCCATCACCAGAACTCTGAGT
TAATTAGGCTGTTGAGTGGCCCAGTTATCAGGATTCCATAAATTTCCTATGTT
GATTTTAAGGTGCAGTCACAATTGAGAGCCATGGATGTAATAGTTTTAGAAT
AGCAACTCTAAGCCATAATGAATAATATAATGGCCTAACACAGTTAAAAATA
TTATTCTTTGTTGTTCTTTTTTCTTGGGTGTATCTCAGTAGGAATGTATGTTAA
ATTACTGTGTTAAATATTTTGTGACATGTTTCCTCTGTGTGGTAATGTCACAA
ACTTGATATACAGTTAGGTTCCATTTATTTCATTTTGCAATTGATTTTGAGGAG
TTTTTTTTTCTAATTTAACTTTATATTTATGTGGAATATTTATGTGTTCCAAAG
TGAAATCTATACATCAAAATATGTTTAAAGTAGCCTGACTTGTATCTCTGTCT
TCTCTACCCTGTTTTCTCCCTCTCCTAGGAGTAATTTTTTGGCTTTGATTTATCC
TTTAATTTTAGTATATGTACTGCTGAAACAAGCACAATTCTTTTCTTTTAAAA
AAATGAAATAAAGTCCCCACTTCTTAGATAAATAGGAGCAAATTATAAAGAC
TTTTCTCCATCTAGATCCCATTTTGGTAGCACATATCTTAGTAACGCCTTCTTC
AAGGACTGGTGAAATTGCGTGTTTACATGGACACACAGAGGAGAACAACAC

ACAGTGGGGCCTATTGGAGGGTGGAAGGTGGGAGGAGGGAGAGGATCAGGA
AAAACAACAAATGGGTTAATGGGTACTAGGCTTGATACCTGGGTGATGAAAT
AATCTGTACAAGGAACCCCCATGACACAAGTTCACCTATGTAACAAACCTCC
ACATGTACATCTGAACTTAAAAGTTTTTTAAAAAGCATTACCGGCCAGGCGC
GGTGGCTCACGCTTGTAATCCCAGCACTTTGGGAGGCTAAGGCAGGTGGATC
ACGAAGTCAGGCGATTGAGACCATCCTGTCCAACATGGTGAAACCCGTCTCT
ACTAAAAGTACAAAAAATTAGCTGGGCATGGTGCTGCGTGCCTGTAGTCCCA
GCTACTCGAGAGGCTGAGTCAGGTGAATTGCTTGAACCCAGGAGGTGGAGAT
TGCAGTGAGCTGAGATCACACCACTGCACTCCAGCCTGGTAACAAAGCAAGA
CTTCGCCAAAAAAAAAAAAAAAAAAAAAAGGAATTACCATTACCTTTATTAAA
TTTCTGAAATCAGATCCGCAATCTGCAATGTTATAAAGATTACATTTGAATCC
TTTCGTGTTGTCTTCATTTGAATCTATACATCATAGCAATTTGATATTGTCATT
ATATTGTCAGTATGATACATTATCCTCAAATTCAGGTATTTGCAACATTACAT
AAAAATAGAGTGTTTATTTGAACCGCGTCTAATCGGGAGAGAGATAAGAAAT
ACCCATCACAGTATCAAGGAGTTTCTTAGCTTAATTACAATATAAACGCCTTT
CCATGTATTGAATTAAGCCAGTGTCAGATCTGTGTGTCTGAGAGTAGAAAATT
ATCAAATACAATTTTAAACTCCATTTGTTTTTGAGACATTTATGAGATTTGGA
GCTAGTTTTAGCTTTAGGCAAGTGGGTATAGAGGAAGGTGCCTGATAAATGA
TATAGCTTCCTTCTCTGATGATTTTTGAGGATTATTTTTATTGTGCATATTTCA
CAATTATATATGCATTCTATTTTTTTCTACAGTCACAGGATTTGATCTGTGGGG
CGCGGTAGTGGCAACGGGGGTGGTCTGCACATTCTACTGCACACTGGTACGT
CCAGGACATATTTCCCTTTTCACTCTACCCACTTGCTTTGCAAAATTGAAAAT
TCCAGTTGTTGTATACCGCAATCTTGTTTGTCCACACTTACTCTCCTATTCCCA
ACCTCCTGCAGTGTAACTTTTTTTGTGTGAGAAATAATCTGTAGTATAATTTG
ATCCTTTGTAGAAAATGGAGCATAACTGAAATTTTTTCTTTTATCTGATCCATT
GCAATGGTTCCTAATCTTGTCTCTTGATTCATCTGTACTGTTGCCAAGTTATTA
TATTTGTGCTCATACTATTCCTCAGCTTAAAAACCTATCATAGGTCATAGGCA
CAGTGGCTCCCACCTGTAATACCAGCAATTTGGGAGGCCGAGGTGGAAGGAT
CACTGGAGCCCAGGAATTTGAGACCAGCCTGGGCAACAAAGTGAGACTCTGT
CTCTACAAAAAAACAAAAAAAAATTTTTTAAAATTAGGAGGCATGGTGGCATA
CACCTGTAGTCCCAGCTACTCAGGAGGCTGAGGTGGGAGGATCGCCTGAGCC
CAGGAATTTGAGGCTGCAGTGAGCTGTGATTGCACCACCGCACTCCCGCTTG
GGCAGCAAAGAAAACCCTGTCTCAAAACAAAAACAAATCTGTGATAACTTCC
CCTTGCTTATAGGTAGGGGTGACAAAATATGCAAAATCAGTATGGCATGGCA
CTGATATTATTATGTGTGCTGGTCCGAGGCAGTGAGAAAATCTTATGGAATCC
TTGACACTCTATGTCTTAACCCTTTAGGTGCTGGATCTTGTAGAAGATGGGGC
CCTCAGCCAGGTGCTGGTTCTGCTTAGGTCATAGAGAAACTTGGGAAATTTCT
TGTGGTGGCTAGTACAGATCACTTATAGTATATAACAGTGGAAATGACCATC
TGAGTGAAAACCAATTAAATATCACCTCTGCCTGTAGGATAAAATTCATATTA
TTTACTAGTAATTTGACTCGGATTTTTCTTCCCAACAATCTCCCGACACCACTC
CACTTTATTCTGTGGAATGTATCTGGCCCCATTCAGCATGCCATGTAGTTTCA
CTCCTTTATGTTTTTGCCCATGCTGTTTTCTCTGCTGGGCGGAACCTCCATCTT
TGCCTCTGCTTATGTTTGAAGACTTGATTTAAATATATCCTTCTGTATGAAGAT
TCTCCCAACCTGACTGTTTCCCTTGAGTGCACTTCTTAGCCAACATTCATGTTT
ATATCATTTTTCTCCAGTATAGTGTATAAGTGTGTAGATGCTTAATAATGTCT
AATAGAAGCATTAATCCTAATATACTTTCCCCTCAAAGGGTGGTCTTAAAGCA

GTTATCTGGACAGATGTTTTTCAAATTGGGATCATGGTGGCTGGATTTGCATC
CGTGATTATACAGGCTGTGGTGATGCAAGGTGGAATCAGCACTATTTTAAAT
GATGCCTATGATGGTGGAAGATTAAATTTCTGGAAGTAAGTGTCTAGTACTTG
GGTAACTGAACACATCTTTTGTATTCTATAAAAATAATCTCTTTATTGAAATA
GTAGATTTACATTAAAAAACAAGCCAACAAATTGCTAAGGATGTGGTAGAGC
AAATTGAAGCAGAGAAGTAAATACGTAAGGAGCCTCCCTCTGTTCTTTAAGG
GATTAAACCTGTCAGATGGTACTTAGCTACATGGTGCTTAGAGCAATGTTTCC
TTCTGAGAAGGGACTTAAAGCAAAAAAAAGTATTTTCTTCCAGGTATTAAAGT
CCAGAATAGGTTGAAAAGTGGGACAGGGTGATAAGGAAAGAGACAGTGGAA
AGTTAAGAAAAGGCAGCTTCTGGCCAGGCACAGTGGCTCACACCTGTAATCC
CAGCACTTTGGGAGGCCAAGGTGGGTGGATCACCTGAGGCCAGGAGTTCGAG
ACCAGCCTGGCCAACATGGCGAAACCCCATCTCTACTAAAAATACAAAAAAT
TAGCCAGGTGTGGTGGCAGGCACCTGTAATCCCAGTTGCTTGGGAAGCTGAG
GCAGGATAATTGCTTGAACCCAGGAGGCAGAGGTTGCAGTGAGCCGAGATCA
CGCCACTGCACTTCAGCCTGTGCAACAGAGTGAGACTCTGTCTCAAAAAAAA
AAAAAAAAGAGAAAAGGTAGCCTCTTAAGAGACAAACACTAGCATATTGGA
TTAGCACCTGCCATAAAAAAAAAAAAAGAGACAGAGACAGACACTGAAAGAC
AGTGATTTCATTTTGACAATTCTTTGTTTTAAGCAACTTTGAGAGTTTCTCTTT
TGATATTGCCCTGGCAATCTATAGTATATTAACATAGTAGCTTATGATTATGT
ATTATAATTCTATGTATGTGTATGTATTATATAATGAACATTATCTAATATGA
AGTATTATTTCTAGAGAATTGACTAGGAAAGTTACAGTCTATGCTTCAAATGA
CAGTACAGCACATGATGTAACATGTTTTAAGCAGTTATTCAGTTTTCCTAAAG
AAAGAACAATGAAGAGACTAGATATTTCATCCCAAGTATATCACACAATTCA
AAAGAATATTGAAAATGCCTTCCGTTTTGGATCAATAGTGTTGTCCTTTTGCA
ATATGGAAAGGGACAACCCATGTTGTCTTGAACTAGCCTATCTTCTCTTTGAG
ATCGCAGCCCCCGCTTTAACATAGGCAGTTTGAAGAAAAAAAAACCCATTTTGC
ACTTGGTGGCTCTTTTCTGGTCTTCTGAAAATAAGCACCAAAGTTTGAGAAAA
AGCTTTTTCAGAAATTGGACAGGGTCAATGTGTTAATTTACAGGGATAAATTT
TAGTGAATCAACTTTGACTATTTTCAATATTTCTTTCCTTCTTTTAGCTCAAGA
TTCAGATTCTAAGGAAAAGAGGCTCTAGTTGCCTAGCTTGGACCTTGGTTCCA
CCCCTTGGCCAGGGCAGAACAAGATATTTTGACTGATCGTCAATTGAGACTA
TTTAATGGAGAAATGGTAGTTTCCCCAAAGCAAACCTGGGTTGCTTTTACCAG
AACAGGGAGAGGGGAAGTCCAGGAAGTGGAAACAACCAATGACTACTCTTG
ATTGCTCCCAAATCTCTTCCTGGTAGGCCAGGCTGAGGAGAGAGGGTATGGA
ACCAATCATTTTTCTGCAAGATAGCTGTCACTTTATATGAAGGATACATATAT
TGGAGCACAGAAGTGATAGCTTATACACATTAGTAAGAGATTTTTAAAAAAA
GATTAAACATTTTTATGACCTTAGTTTTGAAAGTCATTAAGTAACAATAAAAA
GCCTTATTTGTGTTTCATACTTTTCAAGAGTATCCCTTGTTATTCTGATATGTT
TTGAGAATATGATAGACAAAGTATTGCTCAAGTTAAGATAAATGAAGAGAAG
AAGGATATTGTATTGGTATTTTCTGTGTCGTCTGCCATTGGCCTATGTATTCTG
TATGGCCTAAACCACTAAAGTGAGTGTTTTAAATCTTGTCATACTTCCCTCAG
ATAGCTTACTGACTCCTTTTGGCTCTTCAGTTAGCTAATATAGTAGCTTCTCTC
TCTTGGGGAGAAAGGGTCAACAGTCTTGAAACTCTTACACTTTGATATGAAA
TGTTAGACATGAAATAGAGGCTCTCACATTTCCACAAGAATGCCAGAATACA
CATACAATTAGAGCACTTAACTGATCTCAAATATAAGATTTGAAATGAATTTG
CAAAATTTCATAATTTTAAGGAAGTCTCCATGAAAGCTAACTTTGCAGAAAG

TTTTCTAGCTCATATTGTATATCAGAGATGAACAAAATCATTCCTTCCAGTAA
AAAAGATTAAAAGATCTTTCAAACATGGAAGTTGAGCTTTCCCTGTGACAAT
GTTTTGGACTTATGTACATGATGTCATAAAGTGGCTTTAAACTATTAGTATTT
AGCTTGCACGCACAGCATTTTAATAAAGCATGATTACAATGACATGCAGTTTT
TAGAAAATGCAAAACTTGAAACTGCATTAATTGACTGATTTGTTGAACGTGA
ATATAGATTAGAACTATGATTACATGTGCTGGGAGAGGAATTCTATACACAA
AATGTGTTGATTTCGTGTTCATTTTAGTGCCATCTCGTCTGCTGTGGTGCACTT
AGGAATGTTTTTCTCTTTCTCTTTCTCTATGTTTCATGTTCTTCTCTGATTATAT
CATGTCAGATTTGTTTAAATGACAGTTTCCTTAGATCACAGGAGAACAAGTTA
TATAAATGCTCAGAAGGATTGCAACATTCATTTCCCCCAGTCTCTTCTAGTAT
TTTCAGATCTCTGAAAATAGTATTTTTTATTTCATTGAACTTAAGTTGGAAAT
ATTTCTACCTTTTAAATTAATTAATAAATTTTTACATATTTATGTATGTGTTGT
GTATATAAATGTGCATTGCTTGCTTCCAAATATAGTCTAAGAGAATTTTTTTT
AGTCTCTGCTAATAGTTCAGATTTGTTTTCTGTTTCTTGTCATTTAGTTTTAAT
CCTAACCCTTTGCAAAGACACACCTTCTGGACAATTATTATAGGAGGGACCTT
CACATGGACCAGCATCTACGGTGTCAACCAATCCCAGGTGCAGAGATATATT
TCTTGTAAAAGCAGATTCCAGGCAAAACTGTAAGTCACACACCATGGTATAT
GAATCATTAATAGCCATCAGTTGTCTTTATGGAAACTCTTTCATAAGTCACGT
TTAGCTCCTTTATGTCTTTTGTGTCATGAAATTCCAAGAGATAAATGATATAT
TTGGTTACAAAAGGACCAAGAACAAATTGTTACTGTATGTTTTAAATCAGCTA
TTTGAAAATATTTATAGGAATTATTAATGAAAACAAATCAGCATTTATTGTGC
ATCTGTTTGCATAAGAACCCATGGGGGATAAATATATATGGCTTTTGTCTGAA
GAAAATTAACCTCAAATTAGGAAAATACATGCAAGTATGTGAAAAAATAAA
ATAATACTGGAGATATGTAACAGTTTCATGAAATGATATAGCAATAGAGAAA
ATGTTCCAGAGTACATATAGTGCTTTATATTTCATTTTCAGACACAGTCATGG
ATTTATAACCCTGCTCTAATATTTGCTATTTGAGTGATCCTGGGCACATTCTGC
AGTCTCTTTGAGCTGCAGTTTCTAAACTTGTAAAAGGAGCATAAGAAATACT
ATACACCTCATGGGAGGTGTCGTAAAATACTGTGTTTGGCATACGCATGGTA
AAGGCCTAGTATATGTAAGTTCCTTTTTCCACTTCAACTGATGTGATGTGAAG
GTGGAGGATGGATGAGAGATTCCTATTGAACTGGCAGAATGAATATGAAGAA
TGATCATAATTTGGGCAGGTGGAGAAGAGTAGGAAGGCATTCTAGGTAGGTG
GAGTGATTTGGATGAACGCACAGAGGTAGGAAGGACAGCATGGTCCAGAAC
TCTCAAGCCCATCTTGGCTTGAGCAAAGAGTGTGGAATGGTGGCTGGTTCCTT
CTCTGGAGGCCTCACATCACTGCAGTGCTCACCACTCCCTGTGCTGATAAATA
GGCTGACTCTGCTGACATGCTTTTAGGGGCTTTAAAGCTCTAAGGAGATTGAA
TAAGAGTCCAGGGTGGATAGGGTGAGGGAAGGAGGACAATTCTATTTCTATT
TAAAGGAGAAAATGAAAATATTGTGATGTCATATGTCAGAACTCAATATATT
GAGAGTAAATTGGTTTAGAGATTACCTAAATCTTTGAAGAACTCACAGTGAA
AAATCTATATGAATATGGATATAATAACCAGATATGGGATCCCCCCAAAATG
AACTTAGGAATTCTGGTATTTTACGGAATATTGGAGGAATGTAATTTATAGAT
AGAGTAGTCTATTGTACTTTTAAATCTATACTAGAGCTGATACCTCTCCCAGA
GTAATTGTGGACAAGCTCTTCTTCTCTTCCCTTGGGAACATGAAACCCAAAAA
GCCCGACTTGATAACTTAAATAGCAGCATTAGAGCTTTCTTGATAAAAATCA
ATTCCCATAAGCATAGTGGATCTCCTGCATGAAGCAGACTCTCGAAACCCAG
GGGACTCATTTTCGTCTGCCTTCTCACACTTACGCTATGGTAAGAATGAATCC
CTGCTAAAAAACCAAGACCTACTGACATTGATTGACAAAAGACAGTATGAGA

TGTTGATAAGTGGTCAATCTGAATAGCATCAAAGTGAAATAAAACAATATAT
TAAACTCATATATAAGAGACAGGAGGTGTTTGGACAAGAGAGTATCTGGGCA
TAATTTTCTGGTGATTTGGAGATGAGCTAGCAATAGCAAAACATACTGCAAT
GTTAGTCAATTCAAGGGAGAAGGATAAATCTGATAAACCCACCTTTGATTCTT
CCGGAAGCAGACTCTGAGGTTGAGGTTAGCATGGAAGAGTTTATCAGAGAAT
CGCTATATGTAGAAGGGAAGAAAAGGAAACAGGATTGACGAGAGGGAGACG
CTGGACTGTAATGCAGTCTGAACAGAGGCCTCAAAGGGAGCTCTAGAGCTAG
GGTGGCTCTTCAGACTAGTCCCATGTTAGGGTGAGGGGACTGGGCCTTTAAA
CCTCTGAATCCATTTGTCATTGGATGCAGACTTCCTGCGGGGGTTGGGAAGAA
GGAGGCATGACTTCAGGGATGGTACCTCTTTTCCACCTCGGGCAGCTCATCAC
TGTCCACTACAGCACCAGTGGAAAAAATATGTAGCCATCTTAGCAAGAGAAA
TGTTTACTATTCAACTGATTATTAAGACATAGGATTCAATAACACTAACACTA
ATATCAATAACTAGTATTTAACAGGGGTTTATTATGTGTAAGTACCATGCTAT
ATGATAATATACTGTTTCATTTAGTTTTATAACTTTGTGACATAGGCATTGCTA
TCCTACACTTCAGTGAAGAAACTGAAGATCAGAGAGGTTGAATTACTGGCCC
AGGGTCACTTATGGTACAGCCAGGATTTATGCTCAGGACTGGCTCCAGTGCT
GTGTGTAAACCTTTATTCTCTACTGGAGCACACACAGTATATTACAGTGCTGA
ATATTGTTTGAGAAGATACCTGTCCAAGGAATACAGATTTGCATTCCTACTTA
ATGTGTGGTCTTATAAACAACATTTAAACAAGTTTCATGAGTTACTGTGTAAA
TGTTAACAATGTTTAGCAGTTTACAATTGCATTACTTTTAAAAGGAAAATGAG
TAATAGTTAATGCTCGATTGACTATTAAAATCTTTATTTCATGACAAGAAGAC
CTGAAGTAGTATAACTAGGTACCTTTATAAAGCTAACAATGCCCCTGGAGCT
CCGACCACGAGCATACACTCTTTCATGGGGAAATCCCATCCATTCAAATATTT
CCAAATACACTGTCGATTGATTGGTATCATAAGCAGAATATTAGGCTAGAAT
AAAATAAGTAGAGTTTTCGATAATCAAAAGATAATGTACATTTATTGAGTCT
AATCATGAAGGTCTCTTTTGATCATGTAACAGGTTTCCTAATCTTTGGGACAG
AAATTACAGTTGTCTTGCTGGTTTGTTTTCACTTCTTTAGGTCTCTCTACATCA
ATCTTGTGGGACTCTGGGCAATCCTCACATGCTCAGTGTTTTGTGGGCTCGCC
CTATATTCCAGGTACCATGACTGTGATCCTTGGACAGCCAAGAAAGTGTCTGC
ACCAGACCAGGTTCAGTACCATGTCTTTCTTACAGGTGTATTAATAATATTCA
AAAAGCTTATTAGTTGAGAGGAAAGAGCATTCATATTCTTGTAGAGAAACGG
AAAGTGGACATGCCATCATCATCTTACATTTCTATAAAACTTTGTTAAGAATT
TATTTTAGCAGATATAGCAAGAATGAAGAGTACTGCATCTAAAATGAAAATA
TGGAAATACCAGGAAAAAATCAAGCAGAGTGTTAAATAGGAATAGCTCAAG
GTTGGGAGTAGGAATGTGAACACAGTTTTTTGTTGTTGTTGTTGTTTGTTTGTT
TGTTTTTTAAACTACACAAACCATTGTACTATAGAACATTTTGTGGGTGTTTG
TTAGAGCATTCATTTATAAGCAATTTATCTTCCAATTTTTTAAATGATGGAGA
GATGGATCCAATAAGTGAATATTTACTGAGTACCAACTATGTGGTGAGATTCT
GTGCCAAGAGCTTAACATGCTTCATTTAATTCTCACAACCCTGCAATCCTGAT
TTTACAGATGAGGAAATGCTTCTCAGAGAGGTTATATAACTTACCCAAGGTC
ACATTGCTAATAAGTAAGAATAACTATGAGCAGTTATCAAATACCCACCATG
GGCCAGGCACTACCATAATGCTTCATATAAATTTCAACTTTTAATCTTCACAG
TCACCCTCTGATGTAGGTTCTATGATTATCTGCCTTTCAAAGATGAGAGAACT
CAGGCCTAGAGAAGTTAAGTGAATAACTAGCTAGTAACCTGGAGCCAGGATT
TAAAACCAAGCAAGCTGCCACCAGAGTCCTAACTTTGAACCTCTGTGCCATTT
ATTGTCTTTCAAAAAGGGGAACTAGAATTCAAACACAGGATTGCCCAACCCT

AAAGCCTGAGTTCTTGCCAAAATTATTTTCTAAGACTCACTTGCAGAACAAGT
TCCTAGGGGATTCATTAGAATGAAAATAGATTGAATTTCTGTTGATGCAAATG
CATCTCATGCTCCCAGAAAAATACAACTTGGTGGGCCGAGAAAATAAAAACA
CCCTGGAGCTGTTTCTCAACCCTATCTTAACTTATGCCCTCCTTTTGATAAACA
CACCCTCTCTGTTCTCTCAGAGAAGGTAAGAATCACGTTTACTATGAAGATGG
AGGCACACTTCTGTTACATCCCCTCTAATAAAGAATATTTTGTGACTGTTATC
AATCTGTATTGTCTGTAGTTTGTATCATGAAGACAATAACGACTTTAAAAAAA
AGTTTGTAATATACTTTGCCTTTACCCTGGGCCAAAAAAAAAAAAAAAAAAA
ATCCCTACGGCTTCCTACCTTTGAGACATCTTGTAGAATACATTCAGGGTGTC
TTGCTTGCATACGCTTAGAGAGTCCGTGAAGATTTCTCCCCAACTGATATATT
TCCAGGACTGTGTTAGTTAACAAGTTAATTCAATTAACACTTCACTGGGGATA
CCATGCAAATAAGACAGTGGAAGATCGTGTCAAAACCTTATCCTCGCTCAGT
CGCGGTGGCTCACGCCTGTAATCTTAACACTTGGGAAGGCTGAGGCAGGCAA
ATCACTTGAGGTCAGGAGTTCGAGACCAGCCTGGTCAACATGGTAAAATCCG
TCTCTATTAAAAATACAAAAATTAGCTGGGCGTGGTGGTGGGCTCCCATAAT
CCCAGCTACACAGGAGGCTGAGGCAGGAGAATTGCTTGAACCTGGGAGGTG
GAGATTGCAGTGAGCCGAGATCGAGCCACTGCACTCCAACCTGGGTGAGAGA
GGGAGACTGCCTTAAAACAAACAAACAAAAAATACAAAAAAACCCCTTATC
CTCAAGAGAGCAAGGATATGTTTCTCTATGCTCTGCTGGCATTTGCCCAGAGG
AAGCAGGCCAGTTTCTAGATATAGTTTTACTATTTTCCTCTCCATCAATCCTAT
TTCATTGGTTCTACCTATACTTGAGGGGGCCACTCAGATACGAAATAATTACA
CCACAGATTTTTGAGAAAGTTGAAGATAGGATAGAGGATTTTGGACCTTTTTA
TAACACTATAGATGGAACCTACTTTTGCTATTTGTGGGACATTGTTTAAAGTT
AAACCTTTATGATATCTTGGAAAAATTGGGTCTTGATCTCATATAGGGATAAA
TGCTATTCCGTTTTTCTGAGAATAAAGATTGAGTAAGCTTTGGAAAAGTGGAG
AACACAGTCCTAAAAGAACTGAAACAATCCTAATGTTGAAACATTTCTTTTCA
ACAGTGGAGGAAGTTCTTCCATATCCCATGAGCACACTATTGTTAAATGAAA
TTGAAGAGGCTATGGAAGCCGATAAAATAGGACATGTCATCTACTCTGTACT
GTGGGAGAAGTAATCAATAAGGTTTTAGTGCAAATGAGAGGACACTCTTCGA
GAAAATTGTCCACTTAGGACTCTTTTGATTCGGAAACTGATTTTGTAGAAAAA
CTGCCATGCAACAGAGTCCTGAAGTCACACACTTGATTATCCTAATTTGATAT
TTATTTTTTAAAATATAAGTTATCAAAAAGCAAGTTAGTATCAGGAAGTTTTT
TAACAGAAGCAAGTTTAAGGGATTTCCTGAAGTCATTCTCCACCCATCATTAT
GTCTCGTACCTGATGCACCTAAAATTACATCTTCTGTCCTGGGCAGTGACTGA
AGTTCACAAAATGGCCTTGAGTCATCAAGTAAAGTTAAGTGGATGCTGCTTA
CTTAAGCACAGAGGTGTGTCAAATATTTCCTTAAAGACAACTTATTATTCTCA
AGAATAGCAACTACTTTGTTTTGAGCATTTATTATATTCCATATACTATACTAT
GAGTTTTACATAAATTATCTAAATTAACTTTTACAGCAACTCTATGAGGTTAG
TATTATTACAACTATTTTATAGATGAGGAAACTGAGGCTCAGAACTTCAGTTA
CAAAAGCCGTATCTGTGTGAATCCAAAGCCTCTTTACTTAACTACTGTGCTAT
TCGTTTCTCTAAGTGTTAGTAGTTAAACAGTTTAATTTTAGGTATTTGAAAATT
TCATTTGTGTGGAATAACTCCTTTCAGTTCCCGAAGGAGACAAGACAAATGA
TAAACTAGGCATTCATTAATTTTATTCAGTAGCGAAAGTACTTGGAAATAAAT
TTTGGAATTTTTCAGCTCATGCCTTATTTGGTACTGGACATTCTGCAAGATTAT
CCAGGACTTCCTGGACTTTTTGTGGCCTGTGCTTACAGTGGGACATTAAGGTA
TGAACTATGACTCTAATAACATATGATTTCCCTGTTGGGATCTTTCTTTTATTA

TTAGATACTTTAGCGGGATAATGTGGAGTTCTGGGCACAGACACAGATGGCC
CAAAGTAGGCAATCTGCTGTATGTGTCCTCTCTGTGGTCTGACTTATCTCTGT
GGACTAGCGTCCTGTTGTCATCGGCTCCACTTTTAACTAGGTCCCTGCTCTCT
AATACACCACCATACATCCTATGATGTGATACCTATGGTCACAATAATGACG
ATCGGTTGATGGTAGCTTTCACTGGTGATCAAAAATGGGTGCCACAGTCTTGA
TTGAAAACACACATGGGGCTGAAGCGTGGTCAACTGGAAAATTAGAATGAA
ATCTTCCATTTACATGTTGAATAATATATACTGCCCAAAGAATCTTACATTTT
GTGATCTATCATTGCCCTTTCTCCTTGCGTTTGTTCCAGAGAATTGTTATTATC
AACATGTACAGTGTGTGTTAGTGGGGATTCAGGAAATTAATATTGTTGATATT
TACAGCACATGGTAGGGAGGACTTATGACAGTCCTTTCTATGCACAAAGAAA
AATACATTTTAAAGTTGTTATGCATAGGAATACGGAGTAATCTATGTAGACTC
TTTTAGACACTGAGGATTAACAGCAAGTGGAAGCAACATGAACATATCCTTT
CTCTTTTACTGTCAAGCCTGTAGATATTGCCTGAATATCATTTTTGGATGATA
ACAGTTTCAAGAAAGACAGTGCTGTGATTATTAAAAATAGCATAGTAGAGCC
TGGTGTAGTGGCTCACGCTTGTAATCCCAGCACTTTGGGAGGTCAAGGCGGG
TGGATCATGAGGTCAGGAGATCGAGACCATCCTGGCTAACACAGTGAAACCC
TGTCTCTACTAAAAATACAAAAAATTAGCCAGACTTGTTGGCGGGCGCCTGT
AGTCCCAGCTACTTGGGAGGCTGAGGCAGGAGAATGGCGTGAACCCGGGAG
GCAGAGCTTGCAGTGAGTCGAGATCATGCCATTGCACTCCAGCCTGGGCGAC
AGAGCGAGACTCCGTCTCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAG
CATAGTAGAAATCAGGTTATATTTTAGAAGTGACAATTTGTTTCCCCCTTTCT
CATTCCCACTTAAAGTATTATTAAAGGTAAAATATTATAAATCAAAAGTTTTA
TTTTCCCTTACAGCACAGTGTCCTCCAGTATTAATGCCTTAGCAGCAGTAACT
GTGGAAGATCTAATCAAACCTTACTTCAGATCGCTCTCAGAAAGGTCTCTGTC
TTGGATTTCCCAAGGAATGAGTAAGTTTCTGTTTTCATAATTCCATTTTAGCTC
AGAGAGATGATTTTTTGAGACACAAAAATTTCTTTTCCACTGAAGCTACAGA
GGAAGGACCTCTGAATATGACTGGATATCCACATATGTATGCCTATGACAAT
GCAGATTTTTAAAAAATGTTTGTATCAATGTTTAATGTTACCATATCTGTAAT
CAAGATTTGGGAGACACTTCAAACAATTAATTGTCAGTGAACCACAAAGGAC
AATTTCCAGGGGACCGATTTAGCTGTCCTTTTTTCCTGTGCCTTCATCCTATCC
TAAATTTGTGTTAAAATTCTTAGCCACACACACAAAAAATTCAACTATTTTCC
TTTTCAACTGTAGTCCACAGTTCTAAGAAATATCCCTAGTTTGTAACCAAGAG
CCACACTTTTTCTGTTACCTAAGAAGGCACTGTCAGTTTCAGTTATGTTGTCTT
CCCATAAACACTTCCCAAATGTTTACAGAGGATTAGATTAAATTAGATTAATA
GATTAGATGAATTGAAGATAAAGAAACAGAGTGTTATGAATTTTGACTCGTC
TTAGTTGTCTGTTCTGTCTGCTTATGTACACTGTCCTGGAGATGAACTATAAA
TTTGTGCAAGAAATTCTCAACTTCTGTTCTGTTCAATCGTAGAGCCTCATTAG
GGGTTAAATACCAGCTTGAATAGAGTGGTTTAAGCATCTTCAGTTCCCAGATG
TCTCAAATGTAATATCCAACTCAAAGAAATCTTGCCAATGTACTGCATTCTTC
ATTTTGACACCTTGAAATGCATGACTAACAAATTCCTTTTCGAGAAAGAATCT
TTAACCTCAACACAATAACATCAATAGACTGTCAAAGAAATAGTAAATTATA
CCCCATTAGTAGCCATAATTCTATGTAAAATTGCCACAACTGTAGCCTGGAAT
GAGTCCTTAATTTATATCCTTCAGTATTCCCTAAATTTAAATAGCAATGCATC
ATTTTATTATGTACCCAATTTTAATCCTGGAACATAATTTCAGAATGATGCAA
TGCTTCCTAAAGGGTTTTTAGACTGGTTCATTATGGGATTATTGAAGCTGTGT
GGTTTAGTAGAATGAATAGATGAATAGATGCTTTTGAAGCCAGACAAACCCA

GTCTTGCTTTGTTAATTTACTAGCTGTTAACCTTGAGAAAATCCAGTTACCCA
TACAAGCCTCTATTTCTCTATATAGGAAATGGGCACACAAATACTTTGCAGAA
TTGGGTTTTGAGATAATGCATATAAAACGGGAGGCACTGGCTGGTGCTCATG
GAATGATGTTGATGAATGACAAATCAACTCTGCTAATATAGACGAGACTTTC
ATTCATTTTAAAGGGTACTATGAAATCACAATCTTGAAGTATATGCAGTAAG
GACATAGAGAAGAATGTGGCAGATGATTTGAATGTGTTTCCAATAACTTTGA
ATTCCCAAAATTATAATGGTGACTATTCATAACCTATCAAATGGAGAATCAA
ACCAAATATATTGAAAATATTATTTACATTAAAATTGAGAGGCGTAAAAATC
ATAGCAATGAGAACTCTTCAAAAAAATCAATTTAAAATTTTTCATTCAGATGC
ATTATTTACTCAATACGTTTAACAAAGGTATGAATGCTAAAATAAAATAGAA
ATAACATAAGGAATACAAAAATTTTATGATGGACAAATTTCCCATTCATTTTT
TTTCTGTATATTGAGTTTTGCATTTGGGAAATTATATCAACTGATTTCAGGTG
ACTTTTGCTGAATGGAAGTATTAAAACAAAGGGGGTTTTTTTGCATGATATTT
CCTATAATTTAAAAGGTAATACATGTACCTGGAGAAAATTTAGAAAATACTG
AAAGTCAATAGAAAAAAATATTCCACAATCTGACCACCAGTGGTGGTTTGAA
AACTTTTATAAAGAAGTTTGGGCTGAGTGTGGTGGCTCACACCTGTAATCCTA
GCACTTTGGGAGGCCAAGGTGGGAGGATCACTTGAGCCCAGGAGTTTGAGGC
CAGCCTGGGCAACATAGCGACACCCCGTCTCTATGAAAAAAAAAATTTTTTTA
AAGAAGTCTGGGGAAAACAACTTAGCATTAGGGCAGATGTGCTACTTATCCA
GAAGTTGCCTTTCTTTGCTAGTTTAATAGGAAGGGCTTGAGGATACTGATGGA
GATTATGAGGGGGCTAAAAGTCGTCCAACACCCCATAGTGTCCATTGCCACT
TCCCAAGGGAAATGAATGCTTAAAGTCAGAAGAGTCTAATTTCTGTTTATTAC
TCCTTCTCTCACCTTGTACAGAGCAGAGCTGAATAGTATTCTATTTTTGGCAA
GCTGAAAACAGAGACCTGAGCCTTTCTTTATATACAAATGTTTATGGATGATT
AGATTAATAACACAATATAGTTCTTAGTTTTAAATACCTATAGTTTATTCCAG
GAACTCTTTACTTATATAACCTACTGTTGTAACTAATCCTGGGACACAATGTA
AGGGCTTCGTCCTCTTGAAACACTGCTGATCCTAGAGGAAAATAGCCATTTCC
TTTATTCACTGGCTCTGATGTGTGTGGCCATTCTTCACCACAGTCATATTATCC
ACTTTGAATCAAAGGTGTGGTGGATTATTCTATTGAGAATTCTAATTCTCTGG
GTGTGGATTTTACACTGGCTTTTATGTTGTCCATTTAGGTGTGGTGTATGGAG
CCCTGTGTATTGGAATGGCTGCGCTGGCGTCACTTATGGGAGCTTTGTTGCAG
GTGAGAGCTGGCCCCTGGAGGTTTAAGTCATAAATCACTAAATCTTTTTTCAA
TGTTGATGTGACCATCCTTCCAGACTTCTCTCGATATATATCGACACCTGGAC
ATATCAAGTGGCAGGGATGACTACACTTTTTAATTTTTTTTAATTAAACTTTGT
GTTTTGAGATAATTGTGGATTTACATGCAATTGTGAGATATAATACAGAGAG
ATCTCATATACTCTTTACTCAGTTTCCCTCAGTGGTAACATCTTGCAGTGGTA
ACATCTTGATAGTACAATATCAAACTCATATATTGACATTGATATAGCCAAGA
TACAAAACATTTCTATCACTACAAGAATCCTTGCTGTTGCCCATTTGTAGCCA
CAACCACTTCCCTTCTGCCCCTACTCCCTCCTTAATCCCTGGCAACAACTAAT
CTGTTTTCCATTTCTATAATTTTACCAGGTCAAGAATGCTACATACATGGAAT
TACATAGAATGTAACCTTTTTCACTTGGCATAATTCCCTGGAGATTCATCCAG
GTTGTTGCGTATGTCAATAATCTGTCCTGTTTTATTATCAGATAGTATTCTCTG
GTAGGGATGTATCACAGTTTGTTTACCTACTCAGCTGATGAAGGACATCTAAA
TTGTTTCCAGTTTTTGAGTATTACAAACAAATCTGTTACAAACATTACATAAA
GGTTTTTGTGTGAGCATAAGTCTTCATTTCCCTGGGATAACTACCCAGGAGTG
CAACTGTCAGGTGACTGCTAAATGTCTACTTTTAAAAGAAACTGCCAAACTAT

TTTCCAGAGCATGTCATTTTTATATCACTAGCATAGACAAATGGCCCAGTTTA
AACCTCATTCTTTCCAGCATTTAGTGGTGTCTTTTTTTTTATATTAGCCATTCT
GATAGGCATATAGTGATATCTCATTGTAGTGTTAATTTGCATTTCCCTAATGG
CTAATGATGTTGAAAATGTTTTTCAGCGACTTATTTTTCATCTATGTATCTTCT
TTCATACATTATCTCATAATGTCTTTTGCTCATGTTCTAATTCAATTGTTTGCT
TTTTTTACTGGTGAGTTTTGAGTGTTCTTTATGTATTCTGTATACTAGCTCTTG
GTCAGATGTGGTTTACAAATATTTTCTTACACTGTAGTTTGTCTTTTTATCCTC
ATAACAGGGTCTGTCAAAGTGCATTTTTTTTTTTTTAGTTTGGATAAAGTCTA
GTTTATCAATTTGTCCTTTCATGGATTGTGTTTCTGGTGTAAAGTCTAAGAACT
TTACCTAGCCCCAGCTTTTGAAGATTTTCTTCTATGTTTCTTTTCAAAGAGTTT
TAGAGTTTTACATTTTATATTTAAGTCTACAATCCCTTTGGAGTTAATTTTGTA
TAAAATGTGAGACTTAGGTTGACATTCTCTTTTCCTCTATGGATGTGCAACCA
GCACCATTTGTTGAAAAGGCTTTCTTCCATTGACCTGCCTTTACACCTTCGTA
AAACGTCCATTAGGCATATTTGTGTGAGTCTATTTCTGAATTCTCTGTTTTCTT
TCATTTATTTATGTGTCTGTACTTCTGCCAATACCACACAGCTTTATAATTTGA
TTATTTTGATTACTGCAGCTTTAAAATAAGTTTCAAGATCAGGTCGATCGATT
CCTCCCACTGTATTCTTATTTTTCGAAATTGTTTTAGCTATTCTAGTTCTTTTGC
CTTTCCATATGAAGTCTAGGATAATCTTGTCTGTATCTACAAAAAAAAATCTTG
CTTAAATATTGATAGCCTGAAAGCTTTTTATCCATTTGAGAAGAAATGACATC
TTTACCATGTTGAATTTTCTAAAACATGAACATGGTATGTCTCTTCATTTATTT
AGCTTTTCTATGCAAATCGTATTTTTTATGTTGATGTCTGTGTGTTCAATGCTA
AAATGTAGAAATAAAATTGATGTGTTTATATTTATCTTCAATCTTGTGACCTT
GCTGAGCTCACTTATTAGTTCTGATAATTTTTTGCTTCTTTGTTTTATGGTTTA
GTTTGTTTTGTTATATTCCTTGAGATATTCTACATAAACAGTCATGTCATCTTC
GAATGGGGCAGTTTTATTTCTTTCCTTCTGATCTGTATGAATGCCTTTTATTTC
CTTATTGCACTGGCTTCAACTTCCATATCATGTTGAATAGAAGTAGTGAGAGT
GGAAATCCTTACCCAGTTCCCCAATGTGAACAGGAAACTCTCTATTCCTATTC
TCTATTCCTATTTTTTTCTGAGAGTTTTCACCATAAATGGCAGTTGAATTTTTT
CAAATGCTTTTTCTGTAATCAATTTATATGATCATGTGATCTTCTTCTTTAGCC
TGCTTACAGGATGGATTACATTGATTGGTTTTTTAATGCAGAACCAGCCTTGC
ATACCGGGAATAAACCTTGTTTGGTCATGGTGTGTAGTTATTTTTATATATTG
CTGAATTATATGTGCTAATATTTTATTAAGAATTTTTACATCTATGTTCATGAA
GGATATTGATCTGTAGTGGTGTGTGTGTGTGCATGCATGCGTGTGTGTGTGTG
TGTGTGTGTGTGTGTGTGTATACTGCCTTTGGTTTTGATAGCAGGGTTATACT
AGCTTCATAAAATAAATTGGGAAGGATTCTGTTTTCTATTTTCTAGGAGAGAT
TGTCTAAAATTAGTGCTAATTCTTCCTTAATTATTTGGTAGAATTCTCTAGGG
AAAGCATCTGGGCCTGGATTTTTTTTTAGTTTCAAAATTATGAATTTAATTTCC
TTAATAGTTACAGAGCTATTCAAATTATCCATTTCATATTCGATGAATTGTGA
CAATTTATGTTTTTGAGGAATTGTTCCATTTTATCTAAGTTTGCAAATTTATAC
ATGTATAGTTGTTCATAGTAGTTCTGTACTATCCTTTGGCATCTGCAGGACCT
GTAGTGATAGCCCCTGTTCCATTCCTAATATTGGTAATTTGTATCTTATTTTTT
TCAGTCTTGCTACAGGTTTGTCAATTTCATTGATGTTTTCAAAGAACCAGCTT
CTTTTTTTCATTGATTTTTCTGTGTTGTTTTTCTGTTTTCACTCATTGATTTTTA
ACTTTTATCTTTATGATTTCTTTTCTCCTTGCTTTGAGTTTTAATTTGCTGTTTT
CGAGGTGGGATCTTAGATTACTGATTTGAATCTTCTCCTCTTTTCTAATGTGTG
CATTTATTGCTGAAAATTTCCTTCTTGGCACCAGTTTAGCTGTGTTCCACAACT

TTTGACATCTTGTATGTTTGTTTACAGTCAGCTGAATGTATTTTTACGTTTTCT
TTGAGATATTCTTTTGATTTATGGATTATTTAGAAGTATGGTACTTAGTTTGCA
AGTATTCAAAGATTTTCCTGTTATCTTTCTGTTATTGATTACTAGTTTGATTCT
ATTGTGATCAGAGAACACACTCTTTATGATTTCAGTACTTTTAAACTTGTTGA
GGTCTGTTTTATAACCTAGGATATGGTATTTTGGTATATGTTCCATGAGCACT
TGAAAAGAATGTGTATTCTTTTTTTATTAGGTAGAGTGTTCCATAAATATCAA
TTACATTCTGTTTGTTGATGGTTGCTGATTTTCTGTTTGGTTGTTTTATCAATTG
TTGAGAGAGGGGTGGTGAATTCTTCAACTATAATTGTGTATTTGTCTATTTCC
CTTTTAGTTCTATCAGTTTTTGCTTTGCTTCATGCATTGGGTCTTAATGAAGGA
TTAACCTTTTTGTTATTATATAATGTGCCTATCTGTCTCTGAGAAATTTTTTTG
CTGTGAGGTGTATTTATTTTATATTAATATAGTTACTTCTACTTTCCATTGATT
AATGGTTGCATAGTATATTTTTCCATTTTTTCACTTTGAACCACCTATGTTGCT
ATATTAGAAGTGAGTTTTTTTGTAGACAGCACACAGTTGAGTCTTTTTTAAAA
ATCCATCCTGTCAATCTCTGTCTTTTCATTGGTGTACTTAGACCATTCACATTT
AATATAATTATTGGTATGTTATGTTAGGGCTTAAGAATCAGTTTTTAGTTGTC
TGTTTGGTATTTCTGTTTTTTGTTTTCTGCCTTCCTGTGAGTTGTTCTGTGGTTC
CTTCTTTCTACCATTTCCTTTCTGTTTAGAAAACTTCTTTTAGCCATTCTTTTAG
TGTTCCTTCACCTGAGAATGTCTTGAGTTTCCCCTTTATCCCCAAAGGATATTT
TTGTTGGGTATATGATTCTGAGTTGCTAGTATTTTTTCCCCAGGAGTTAAAAA
AATATTTTTCTACTTTCTTTTTGTCTCTATCATTTCTGATGATAAATCTGCTACT
ATTCTAATTGTTTTCCCTTACAGGTAAGGAGTCATTTCTTTCTGAATGCTCTCA
GGTTTTTTTGTTTTGTTTTGTTTTGTTTTTTCTTTAGTTTTCAGAAGTTTAATTA
TGATGTATCTTGGTGTGTATTTCCTTGTTTTGAGTTTTTGATGTGTCTTGGCAC
GTGTTTCTTTCTTTTCCTATGTTTTTAGGGTTACTCAGCTTCTTGAATCTTGAAT
CAGCTTCTTTCATAAGTCTGAAAGAGATTTATGTCTCTTATCAAATTTGAGAA
GTTCAGACATTAATTCTTTGAGTACTTTTCCACCCCATCTTCTTTCTCTTTTCCT
CTGGGACTCTCATGACATGAATGTTAAATATTTTTGTTATAGTTCTACAGATC
CCTGAAGTTTATTTTATATTTCTTCATTTCATTTCTCTCTATTGTTCAGTTTGGG
TTATTTCTGTTGTTCTGTCATTCAGTTCACCGATCTTTTTCTCTGTTCCTTCTAT
ACTGCTATTGTGCCCATTCATTTAAAATTTTTTATTTCAGTTATTGTATTTGTT
CAGAGTAACAGTTTCATTTGGTTCTTCCTAATATCTTCTTTTTTTCCCTGAGAT
TTTTCTCTCTATATATATTTTTTCCCTCTATTTTTTCATTAGTTTCAAGTGAATT
CTAATTGTTAAAGCATTTTTTAAATCATGGCTGTTTTAAATTTTTTATCAAATA
ATTCTGTTTCTGTAATCTTTATATTTAGTTTGAGATCTTCCTGGTTCTTTCTATG
AAAAGTTAATTTCTGTTGAAACCTGGGCATTTCAATATTATGTGATGAAACTT
TGAATCTTACCTTCTGTTTTAGCCAGCTTTCTCTGACACTACGTCAGTAGCATA
AGGGAGGGTGCCGCCTCATTACTTCAGGTGGAGGTAGAAGTCCAGAGGGAG
GAGCTCCTTGTTATGCTGGATAGGGGTGTGTAGAATACTATGAGACATATTGT
GGCTATAAGATTAATGATAGTGCATGAGGCCCACTGAAATATCTTGCAGGGC
TGATACTACATGTCATGTAGGAATTTACAACCCTGGCTCAGGGATTTCCAGGA
AAAAAAAGCCACCTCAGCACAGAAGCAGCTTTCATAAACCTTAGAACAAAGC
TTACTTTTACAATAATAGCTTAAATACCCTTTATGAAAGAAACAGCTGGTAAC
TAACCTGGACTAAATACAGGTATAAGAAAGGGAGAAGGACCCCCAAAGTCT
GACAATGGTCTCTGGATGAAGACTCTCTGGTCAGTTCATGATCTGACCCCCTG
ACTGTATCTGGCCCATGACACCAGCTTATTCTCACTATCCATCTTCTAAGAGT
GCTGCCAGAATAAACCGATTGAGCATTAGATGGTGCCTAAGACTCATCTTTG

ATGTGAAGTGAACAGAAAAGGAGACATCGCCCCTGGGGAAGCTGGTTAACT
AGGTCCACCTACAACCTCTGAACACAACTGGCATTGAGGATAGGATAAGTAA
GTGAGCAAGTAAGTAATGGCACCCATTTCTAAGGAAGGACTGGGAAAGGGT
GACTGGGGACTGGTTCTGATCAAGAAGTCAAATGAAGCCCTCCGGAACACTC
CGTAGTGTGTCTGGTAGTTTTTATTTTACTTTGTGGCCTGTTGCTGCCTTCTAC
ATACTTTGTCTGTGCCTCTGTTGTGAAATTGCCACCATGAAAAACCTGCCAAA
ATTATTCTTAAGACCACTCTGGTCCCCAGTAACTGGGGACAGGCTCTAAGGA
GGAGGAAGCAGAGGTGGAGGATGCCCCGGCTTGAAGAGCTAACCACTGGCT
GGTGTGTCTCTGGCTCCTGCTGTGGGAGGGTTGCTCCTTGCCCCCAAGGATCT
TGGCACTGTGATGCCACAGACCAAGAAGGCAAAGGTGAAGCTTACCTGCTGA
AGGGTTATGTGCTGTTGCCCACCAGCTAGGGACTTAGAAGCATATTTCCCAAT
AGATCCTTGGACTAACACCACCAGTGAAGTCCACAAGAATCATCAAAGGGCA
TGTCGCAGCTGTTCCCAGCCCACCATGCACTCCAGCTGAGACAGTGTGGCTTC
TCAAGGGGTGTGCCCCAGCCTGTGATGTCTGCTGAGTGTTGCTGGCAGCAGC
CCATAGGCAACATGCCCAAGTCACGCGGGCTCACCTCAGAGTCGTGTCATTC
TGCCTGGTCGATCCAGAGCCCTGTGAGCTGTGGAATGATGATCCTGGGTGTG
CTGGCTTACCGCTGGCCACACCACAGGTGTCACCAGTCATCACTTGGGCTGA
GAAAGCTGCTGCCACCTCCACAGCTGGCTGAGATGAGGAAGTGTGTCCTAGA
CACCATATGGTCCATGAAGAGGTTCACCCCTAATGTGGACAGAAATTCAAAG
CCTGCTCAAAGACCTGGTGCAGGAACCCAAAGAGAAAGGCACCACCTGGCTG
TGCTGTGTGTACCTGAGAGGAAGTGCTTTGCAACTCATAGGCCGAGAAATAC
AAATACAGCAGCCAGCGGCTATCTTTAAAGATAATAAGATCATTTCCTTATTT
CAAGACCCTGTTGAGCTTATCCAGGCACAGAGTAGAGTGGGTAGACATAACT
GTCCCTGCAGTCCGCTGCAGTGGCTATTCAGGCCTGGCGAATAGCACCCCCA
GAATGAGCCTCTTTTCATCAAGGGCTGAAGGTGGCTGGTGCACACTGAGGAA
GAAGGGGAGCCACTGTGGCGCTTAGGCATGCTCCAGTGAGTTTGCACAGCCG
CCACTCCCATGGGACAGGGGCACAACTGGGAGACCTTGGAACAGACCACCTT
GCACCCGGATGGTTTACAGATGTTTTTAAGGAGAGCCCATCGTAAGCCAAAC
CAACTCTCCTGGCCCTGATGGCCACAGTCCCTGAAGTTAATTGAGCTCTGTAG
GTCCTCATCACCCTGGAAAATGCTGAAGAATAGCTAAGAATCCCTAAGAAAA
CTTAGGGAACTCACCCTATACCACTCTTCAGGTCCAGGGTCCCTAGCGGGCTG
CCTTCTTTGCACCTTTCAGACAGAACCATCCAACAGTGTTCACAACAGCACCT
CGCTCCCTACCCACCACACAGGACCTTGCAGCCAGAGGTCTCACACTGGGCT
CTGCTCTCCCTGGCCTTCTCACACCCATCACCCACGGGATGAGGCCCAGTCCA
GCTGTTCACAAGGGCAGAGTGACCTCTACTGGCAGATACGCTGAAGGGTGCT
TCCAAATTGGCTCCTGTCTGGCAAGCCTCATGAATTCCTGCAAACTCTCCTTA
ATACCAATGTCGTCCAGTCATCCCACCCACTGACCCACCACCCATCTGGATGG
CTGAGCATCAACTTCCACCTTTGGGAAGATGTCTTCACTGCTAACATTCCTGC
CCTGTGTCCAGGATGACCCACCTGGACTGTCCCACCCACCCTACCTATATCCC
TCCCTGGCAAGACCTTTACTGGTATATTTGCCTAGGGGTCTCTTGGGCCACCA
CTGTTCCTGCAGCCAGTGCCTCCTAGAGACACTTGGTGCCCCAGCTGTGCCTG
ACAGCCATCTGGAATGGACAACACCTTGAACTTTGGCTTCATCGGGCCGTGTT
CAAGATCACGATTTTGGATCTTTGACAGCACCTCAGCACACACAGTGTCCCCA
GACCTAATTGCTTAGTTCTGTGTCCTCCTTATCCTTCTGTTCCATTTTCACCTG
TCATTGCTCTGTAATCAAGCAACTCTAAGCCTGCTACCACTGAGTCCATTCAA
CTCCTTTCACCTGCTCATGTAATCGTGTTTCCAAGCCCCAGTGTGCCCGTGTTC

TTTTGCTGCTACTCCGGGTGGCCTGACTTTCCACATCCAACTGTGGCTTCTCCA
GACCTATTCACCAATTACTTGGCACTGAGGTCACTGACAGCCTCCACCACATT
TTGGCCACCCACTGAGCAGCTAACCTTCCAGCCTGCTGGGCATGCCTTCATTG
GCCTGCCACACATCCATTGCTTCCCGTCCAGACTTGGTCGCTACCTGTCCTCC
ACTCTTCCCCCATCATTGGCGAGTCTGCCCCAAAGCCACAGCCGAACCCCATT
CTCATGATCTCCCGTAACTGCTCAGCTGGACACTAGATCTACTGTATCCTCTC
TGTGATGATACAGTTCATTTCCAGGTTTTTCCAAGGGCGCCACCAACTCCTAC
TCACCGTCCTACAGAAGATCCATTTCTCAAGATTTGACTGACCATGGAGCGCA
CTTTCAAACTTGGGCAGCCTCTCAATTATTGACCCCCTCAGGGTCAGGGGTGG
AGTGTGGTATACAATGAGATATATTATGGCTATAAGATTAATGATAGGCATA
AGGCCACTCAAGCAACTCAGAGGGTCAGTCCTATCTGTCAAACTTAGCAGTA
TGAAGCAACCTGGCCGGGGATTTCCAACCAGATTTCCAGGAGACCAGGTCAC
CTCAGCACGATGCAATTTTCACAAACCTTGGAACAAAGCTTACCCTTACAAG
CATAGCTTAATCTCTCTTTGTGAACAAAACACCTGGTAACTGACCTGGATTGA
ATACAAGTATAAGAAAGGAGGAAGGATCCCCTAAACTCTGAGAAAGGTCTCT
AGATGAAAACCCTCCTGGTCTGTCAGTCATCTAACCTGTGACTAAATCTGGCC
CACGACACCATCCTGCTCCTGCTATTCTTCTGGTAAGAGCACTGCCAGAATAA
AATGCATGAGCATCAGACGGTGTACAAGACTCAACAATGATGCAAAGCGAA
CTCAAAGGAAGAGGCTTCCCTGGGAAGCTGGTTAACTAGGACCACCCGAAAC
ACGCGAGCACCACAGGGTGGAATTCACTGACATCATGGAGGAGGTGGCTTTG
TCACTGCTGGGCCATGGTGAAAGTCCCGATTCTCCACTAGGCCTCCCATGTCA
TTGCCAGCAGGGAGGCAAAGGGTGCCTGGTGACAGTCTGGGGGGATGGAAT
CTAGGCTGGCATGGGTGTGGGTGGGGTGAACAGATTATTCTGTGGTGTTCGG
ATGGAATGGAGCGGTTGGGGTCTAAGAGTTTTCTATCTTCCTAGGCTGCTCCT
TTCCTGGCCCTGTGTCTAGAGAGAGAACAGGATTTTGTGAGGGCTTTTTATCT
TTTTAAAGTTTTTTTGCCTGTGTCTGTTGCCATTTCCAGGCTGGCTGGCTTCTT
CAGTTCTCAGGCTGGGATACATGAAGCAAAAAGAAAACTTAGGGAACTCACC
CTGTACCATTCTTCAGGTCCAGGGTCCCTAGCAGGCTGCCTTCTTTGCACCTTT
CAGAGTCTTCTTATGTTTATTTTATATGGAATGTCAGGATTTTTAATTTTATTA
GTAGGAGGAATAGAGAAACATATGTTTATCTTCTCAGAAGTAGAAATACTTC
ATTAAAAAAAAATTTGTTTATTTACCTGAAGGGTCTTTCACCTCCTGTGGCGTT
TGCTCTTCTGATTAGAGAGCATTTTTGTGTTGGAAAGCTCCCTCCTGTTCCTGT
GGAGCTGTGGGCCTGGAGAGGACAGACCAATCTTTACAACAGTGTCTTCTTC
CTGAACCCTGTGGGCACCTCACTGTCTTCAACACGCAGGCCTTGGCATAGACT
TCTCCCTCTGCCTGCAACACCTTCCTCTCTACCTCGTCCACTGCCTAAGCCTTA
GTTATTCTTCAGTTTTCATTTTAACACATTACTTATTTAAAGAGACTTTTCCTA
TTACCTGGATACATTTGCATTCCCTTCTGCTAAATGTTTCTAGAAGCACCTTGT
ACTTTCTTTATCATCTCCATTATCAAACTGTATTACAGTTATTTATTTAGTGTC
TGTTTTACTGAATAGCCTGTAATTCATGAAGGTGGGGAAGATGACTTCGTTGC
TCAATCCTGTATCAAATATCTGGTTCCTACTGCAGTGCCCAGGACAAAGCAA
GACATTAACAACTTTAGCTTGAATATTAACTAATAATGTAGTTAGGCTCATAA
ATATCCTATTATCTGATGACCCCAAGTGCTCCCTGAAATTTCCTTGGCCACTT
GGGCTGCTCCTCTGGGATCCTGGACCTCCCCTTGACCTAGAAACTTAAAGGCT
AGTACTGGCACAGCAGGGGTCCTCCAGGATCTTTCTTCAGTCCTAAGGCTTGG
GGCCATTCTGGTTGGTTGTTGTTTAGGACATAGCAGTTGTAAAATAGTTTGAC
CCCATCCCTGATTATATGAATGAATGAATTAGATCCAGAAGCTCCAAGCCTG

CCCAGCACTCCTAAACCACAGCCAGTCATACAGAATTTGCTTTCTTCATAGTC
TGCATTAGTGCCTGGAGCTCTTATTGTGGTTGATTCCCCTTTTTCCATATATCA
GCTATGGCCTGATTATATACAATGACATGTTCTAGCAAAGTCATCATAACTCC
AAGCCAGTTCATCCTTGCTTCATCTCTAGCCAAGGTCTTATTTCCTCTTGTAAT
CTTTTTTGGGTGTCTTTGTTATGTCTCCAAAACTGCTCACAGATCACACAGTC
ATCACTTGCTTTGTGGGTCATATGACGTCTTGCAGTGAAGTCTATCTCCCTTCT
GACTGGACTGTGTGTAAACTGACTCAGGATCACATCCTTGAGTCCGCATCCTT
TTAAACACTAGTTCTCCTCTCTACTCAGTACCACATCCTTTTCTTCTTCCATGA
AGTAGGCCTATTTAAAACAACAACGACAACAAGAAGCACTATTTCCTCTAAT
ATTTCCCTGACTTGGTATTTTCTCTTTGATTCATTCATTCTGCTCTATTAAACC
CCTTTACTCTCTGGACAAACCAGTGATTCCCTGCTCTTCTTTTATGGTTGTGTC
CATTCTTAAGAACTGGGCATTAGATTATTGGGAACGTATATCTCTTGGAAGCT
TCATTTAGTTGCTTTGCCTCATGCGATGTTTGCACTGTAAATCTTTTTTACTAT
CATGGATTTGATATCTCTACATCAAAGCAAAGGTTGCAGTTATCTAATTAGAT
TTTAAATCTATTTCTACATGATTTGATCAGTCTTTTTAAAATCGTATTTCTCAT
TCTAGTTTCATCCTTAGGGAATCAAATCAGGAATGGCAAATGGGTTTCATCCT
CCTCTCTGGTCACATAGCAGTACTCGTCCACCGTTATGAGAGGATTTTGAGAT
TATCAAAGCACAAAGGAGTGCTGTGATTTATAAACACCTGCAAGAGCAATGC
TAAGAGGAGAGGTATCATTCATGATTTATACATTCACATAGGCACTACCATTG
TCGCAGACCTCTCTATCTTCTTTGTGTTGTGTGGCCTTTTTTGAGGCTACTTGC
AGAAACAGATGGATCCTTGAGACTGAGATGCAGAAACTTGTAAGTTCTAATG
ACTTCTTTTCCAGTGATAAGGCTATCATGACTGAAACTATGATTTTCAGAAGG
AGGCCGAATACTTTAAGTCATTATCCTGATGAAATGACTTTGAAATATTTGAG
TTTCGATTTGAGATTGCTAATTGCTGACGTTGTATTATTTTTGTAGGCAGCACT
CAGCGTATTTGGTATGGTTGGTGGACCACTTATGGGCCTGTTCGCTTTGGGCA
TTTTGGTTCCCTTTGCCAACTCAATTGTAAGTACAAAGAATGAATATGCTTGA
GGATTACTTTTTGAACTATACTAGCAGCTCTACACTTTTTTCTCAGTTGGTTCC
TTTGAGATTTGTCATTAGCTACTTGTCTTGATGACTTAAATTATTTCTGTTGAC
TTTGGTGGAGTGTACAAGGAAGTACTATGTATGGGGACCCTAAATGTGTGAA
GCTCAATGGAAACTTCCAGAACCATATAGGGCAATTTTAAATATTCATAATAT
AACTAAAGGAGCAACATTTTTATGCACGTACCTACTTGCCCTTCTCAAAAATT
AATAGCTAGGATTTAAAAGAGATCAATAGGCTCAGAGATGAGAGATTTAAGG
GCAGAAAACTTAGGATTTCTGTGAATAGCCATAGCACAGCAAAGAGAAGGA
ACACAATCTTACCACCTTGGCCAGGAATTTTCTACTCCTGACATTTAAAGCTG
TAGCTCCCACAACATGATTTAGCCCCAAAGGGGTGATCATAATTGGGAATAT
TTTCCAGAAAGACGTAATATCTCTGTCCTTCTACCTAGACATATCTGTGCTTA
GAAGTGCTAACTTTTGTTTGGAATAAATGAGCTAGAAATTATTTCTCTGAAAC
CCAGAAGAAAGTCCACTGGTTCAGTCTGGCTATAAGATGTAGTTCAGGAAAA
ACTGATAATGTATGTGCAGTCGCTTCAGATATAGAATAGCCATAGAACTCTG
ACTTCACATTTGGAATTCTATTTTCCTTATAAGGCATTAGGAATTGGTAAGGC
AGATATTAATAAGGACTATTGTGGTTATTATTTATTATCCTCAATGAAATGTC
ATATGAAAAGCTGCTTTTGTAGGAATTAATCACAATGGAAGAGATTTGTTTGT
GTCCACCTCAGAATGTTGAAGCTTGGATTAGATTCTTTGCATGATGAAATGCA
TTAGTTTGTTAAGTATTAGAAAAATGTATTTAAAAATCAACTTTTGATATTTG
GCATTTGTTGCCAGCGTGTATTTCCTCCGGAAGAGCTGTAGCTGACTAAGCTA
ACATGTCCTTTTCTGGGGGCGGTAACGGAAGAATAACATACTATTCTTCCTGC

CATAAACAGTATCTTTTTTTTTTTTAAACCATGTCCAGGTTTTTCACCAGCAGA
TTATCTGCCTTCCTTAAAAGTGGCCTTTCTCACTCCTTTTTCTCTCCCTGCCTC
AATACCTTTCATTCATTTATTCACTTGTTCATTTATTCAATAGATTTTTATGGG
CTGAGCACTATCTTGGGAGGAGAGAGCGTTGAATGTTTCTGCCCTTGTGAGTT
TTATAATTTCGTGCCATGTTGAACTTGGCTTTCTAATATTAATTTTGACATCAA
AGCAAATTGTATTATTTTTCTTTTAAAGGCATTCTTGTTTTACAGGGAGCACTT
GTTGGTCTGATGGCTGGATTTGCCATTTCTCTATGGGTTGGAATTGGAGCTCA
AATATATCCTCCACTTCCTGAGAGAACATTGCCATTGCACCTTGATATCCAAG
GCTGTAACAGCACCTACAATGAGACAAATTTGATTACAACCACAGAAATGCC
ATTTACTACTAGTGTTTTTCAAATATACAATGTTCAAAGGTATTGAATTAAGT
TTTATTACATTATACTTTAAAAAATTTACGCAACAAGTAGAGAACCCCACTTG
CTTTTTGTCTTGCTTACAACACTGTGATTTTGCCTAATTCTGAAATGAGTAAA
ACCCATGTGGTTAGCTATAGTATTTTCTGCAGCGGTAACAAAAAATGTCATAT
TTTCATAATTTCTCTAGAAATTTCTGCCTTGTCTACAGATCTAAGAAGACTTA
AATATTAATGAGAAACTTCTGTTATGTGTAAACTCTCCTAAACACCAGCTCTT
AGCTGCATGAAGAATTATCTTTGTCTTGGAAAAACTTTTAAAATGGAAAGCA
CAATTATAGAATAAATATTGCTTATATAGTCTTAGAAAGATAGATTTTACTGA
CCAAAAGCTACAATTATTTAAACATGTTAAATAACTGCCATTTGTTCAGTTGA
AGATTCCAAATCTTTAAAGCATTAGAAGTGATTGCAGCTGTGATCTTCATGCT
ACAGATTTCAGTCATGCAGTACACTTTGGAGCCTCTAAATGCTGAAGTTGTCT
GATTTAACGTACTGAAATAGTGGGTAGAGGCATGCTTTATTGTACAGTGAAT
GTGAGGTCAAGACTTTCTTTAGTGGATATATAAGTGTCCAGCTTTAAAGCACA
AACCCTGTGAATACGTTCAAGGAATGCAAAGATGATGCCATTGCCCCTAGAG
TATTGCCCCAGTGCCAGCTTATCTGGAATGACATCAATATAGTCATACCTTTG
GGGTCAAGACACGGCATACCCCTCTTAAAATGGACACACTCCTGAGAGAAGG
AACGTGATCATACATATAGCCATTATTAGCAATTTGTGTTCTGGAGAGAATTC
AGCTATGAGAAAGCTAGCCAGTGGGACTTCCAGGGGCGTTCCAGGCAGGTAA
GCTGTGTATGGTAGAGTGGAAAGGTAGTTCCAGAGCCTGAGGCGTATTTGAG
GGAAGTTTTACTAGGTGAGGTGCTGTGGAGAGAGGGGCAAGGACATGTGGA
GAGCAGTCTGGACAGGCATGAAAGCACTAATGGGAGGGGACTGGGAGAGTC
ATCAGAGTCAGAAAGGGAGAAAAGATTAATTGAAGGTTAAGTGAGAAGACA
CAGGATGTTCTTCTTGTTTTATGATTTTGCTTAGCGTTTTCCTAATTATGTAAA
TATTTGCCCTTAAAAGTCTAAGGCAAAATGGTTATTTGGATTCCTGAAACAAA
TAAATTAGCTGAAGTTTATTCTATGCCTGGTACCATAATAGTTTTATATGTTA
GTTTACTTATCAAATATTTCAGCTAATAAACAAGAATTTTTGAGCATGGACTA
TGTACCCAGCACTGTGCTGGCCATTTATTTAGTATCCATGGAAGAAAATCTTC
CTCTCCCAGCCACTGTTTCCTGCCTTCCAGAAAATTCTAATCTAGCAGGAGGG
AATGTGCAAATATACGCAAAACAATAAACTCAGTAAGCATGCTAAGTATAAA
AACTAAGACATAGTAAATAAAAAATACTTGATACATTTTTGTTAAAAGAGTG
AGTACAAAGTTGGAGTGTAAGTTAGCATGAAGACTTCAGAGATATACTACTA
CTTCAAGAATGCATAGGTTTGGCTAGATGAAGGGGAAGATCTGGAAAGTCAT
TCTTGTCACAGAAAATGACCAGAAGGGAAATTGGACAAGAAAATAAATTAA
AACATGCATTGTGGTAAGTAATATTCACAGGATATCATGGGAGCCAGAAGGA
GGAAGTCTGCCTGACTGAAGTCAAGGAAACCTTCTTAGAGGAGAATTGCATG
GGCTGGAGTTAGTCAGGTAGCAGGATTGAGCACAGGGACACAATCTGTATCG
CAAGGAAGGAGGAGTGTGTGTCGCAGAATGAAGAAACGACCGGTGCAAAAG

CATGAAGGACTGAAGAATCAGGGCATCTTAGGAAATCATCCTAGTTCCATAG
GGCTGGAGCACAAGGGTTGTGAGCCACCCAGCTGAAGGTTGAATGTGTGATG
CTTTGGTTGCCTAGATACAAATTGTTTCAATTGCCTAGATTCAGTTTTGATTTT
TTTAAGTTTACAATATTTGATGGTTACATTGGTAGCTTGCTGTGGTTTTAATTT
GTATTTCCCTTAAGATTAATGAAGTCAAATACATTTTCATTTGTTTGTTGGCCA
TTTGGATATTTACTTTTGCGACTTAGTTATTGAAGTCTTTAGTCCATCTGAAAT
GGATTCTCTGTCTTTTTCTTACTTACATAGGAGTTCATGACATATTCTACATAT
GAGATATGTTTTGAAAATGTTTCTCTCCCATTCTATGGGTTACCTTTTTACTCT
TTTAATGATACTTTTTTTGTCAGTCAAAATTTCTTAATTTTAATTTATTTCTGTT
TCTCAGTTTTTTCCTATATGGTTAGTCCTTTATGTGTCTTGTTTAAGAAATCCT
TGTCTACCCTGCAATCATAAAGATATTCCACTAATTTTTTATTTAAAAGCTTTA
TTGTTAACCATTTCCATTGAGGAAGTGTAATCCACTTTGGAATTGATTTCTGT
GGGTGGTGTGATCAAGATTATTTTTCCCCATGTGGATACTCAGTTAACTAAAA
ACCATTTATTGAAAAGACTCCCCTTTTGCCCCATTGAATGCACTGGCGATGAA
AACGATGTTAACCAAGTAACTATTTTTCTGTGGGTCTATTTCTCAATTCTCTAC
TCTGTTCTATTTTGTCTGTCTTTGGCTCAATATCACACTCTCTTAATTTCATTA
GCTTAACTCAATGTGTTCCCCTCCTTTCTGGGTTCTTGTCACCTCAAGTTCTCA
TTGTTTTGTTTTTGTTTTTGTTTGAGACGGAGTCTCAGTCTGTTGCCCAGGCTC
GAGTTCAATGCAACCTCCACCTCTGGGTTCAAGTGATTTTTGTGCCTCAGCCT
CCCAAGTAGCTGGGATTACAGGCATGTGCCACCACACCTGGACAATTTTTGT
ATCTTTAGTAGAGATGGGGTTTCACCAGGTTGGCCAGACTGGTCTTGAACTCC
TGACCTCAGGTTATCCACCTGCCTTGGTCTCCCAAAGTGCTGGGATTATAGGT
GTGTGCCACCACCCCAGCCAAGTTCTCATTGTCTTGATCTCTCTGATAGCTTC
AAAAACCTGCCTCCCTCCCCATGCCCCAGGTTTTATATATTGTTCTCAGCAGA
ATGATTGTTCTAAATCAAGCAGCTATCAAAACTGAAAGTGGAATCCTAGAGC
AATCATTGTAGAAATGATCATAGAATGTTCATGTTACTCTTTCACTTAAAACA
TTGCAATGAGTGCCATTGCCTGTAGAGTAAAATTCCAACTCATGCCTGGTCCC
ACAAAAAGCCCCCTGCCTATCTTTCATCTTGTAATCACTGTCTTGCTCACTGT
GCCTCAGTCAGCTGGTGTTTAGTCTCTGTTTTCCTCAAGAACATTTTGTTTCTC
AGATCTCAGTTCAAATGTCATTTATTCCCAAAGGTTTTCTCTGACCACTTGTA
GTGTAGTTCCTCCCTCTTCACCCTGTTAATAGCACATCATCTTGTTTTACTTTC
CTTACAGCACTTTCTACTAGTGGAAATTATTATATTTATTAATTTGTTTATTGT
ATTTATCTCTCCATTAGAGCATAAACTATATGTTAGCAGGAATCTTATCTACT
TTATTTTTAAAAAATAATTTCAACTTTTATTTTAGATACCTGTGCAGGTTTCTT
ACTTGGGTAAACTCTTGCCTGTCTCCCTTCACCCTCTGGTAGTCCCCAGTGTCT
ATTGTTCCCTTCTTTGTGTCCATAAGTAACCAACATTTAGCTCCTACTTATAAG
TGAGAACATGTGGTATTTGGTTTTCTGTGCCTGCATTAACTTGCTTGGATAAT
GGCCTCCAGCTGCATCCATGTTGCTGAAAAGGACATGATTTTGTTCTTTTTTC
ATGGCTGCATGGTATTCCATGGTGTATATGTACCACATTTTCTTTATCCAGTCC
ACCATTGATGGGCACCTAGGTTGATTGCATGTCTTTGCTGTTGGGAATAGTGC
TGCGATGAACATATAAGTGCACGTGTCTTTTTGGTAGAACAATTTATTTTCCT
TTGGATGTATACCCAGTAGTGGGATTGCTGGGTCAAATGGTTTATCTACCTTA
TTTATGGCTGACTGTCCAATCACAAGAGCAGTGCCTGGGACATCATGCAGAA
TTGAATGAATAAACATGACCTTCAAGTATTTTAGTATATAGTCAAGGTGAAGT
TTTGCTTTCAATTCCTGATCTGATAAGATAAAATTAGGACAGCAGGTTATGAC
ATTAAAAACATTATTAAGTGTTATATTTTGACCTTGCCATGGGAATCACTGGT

ATATACTGGTAGTTGCTAGTGGCATTACATGAAATCTTACCGTTGTGTTTTTG
CTCTCTTAGGACTCCACTGATGGATAACTGGTATTCTTTATCATATCTGTACTT
CAGCACTGTTGGAACTTTGGTAACATTATTAGTGGGGATACTTGTCAGTTTAT
CAACAGGTAACTATCTAAACATTAGTATTGTTGTATTTACCTCTATATCAGTT
TTTACTGTATCTGTTTTATAGCTATTTATGTTATTTTACTATGATCCTATTGCC
AACTAGTTTAATTATTTCTTCTGTTTTACTCATGGGAAATCTGAAGGCCTAGA
CTTGAAAGCAGAAGACTCATTTCTATTTGAATCTTGGTTCTGCCCTTTAGGGG
CTGCATGAACTTGGGCAAGTCACATAGCTACACTGAGCCTCAGCTTTCTTATG
TATAAAATGAGGACAATAATAATCGACCTGCCCATCTCACAAGTTGCTGTAA
GACAAAAATAAGATGCTTGAAAAGAACTTCATAACTGCTATGATGTTAGGGA
TTTTACACATATTTTTATAATAACCCAATATGCTGTGATCTTTACATGGTCCAG
TATTCATTCTTTGTGAACTGTAAAGAAGCATACACAGGTTATATCTAAACATA
GTGTGGAGGCTTTTTTTTCAGATATTTATCATTTTTTAAAAAGTGTGTTCATTT
TTAAAAAATCAAATTGATACACATAGTATATAAATTGCTCTCTGGGATTGCTA
TAGAACGGTTATCTTGGGACTCCATTTTGTCCTCATTCTATGAATTCTCTTAGC
CCCTGTTCCTGTGTTGGGTTTTCTACTGCCTATATTCATTACTTCTCATTCTCAT
TTACTCCCTTAATTGAATGAAACACATCCTATAGCATATTCTCAAGAAAGAGT
AGACAAAAGAAGAGGTTTTGGAAACTTTGTGTGTCTGAAAAAAAAATTATTT
AATCTATCTTATGCTTGATTGATTGTTTGAGTAGAAATTGCTAGGGTTGAAAA
GAACTTTTCTTTTTTCTGAGAGGATGTCTGAATTTTTTTCTAGCTTGCTATGTT
GCTGTTGTGACAGCGCTCTGCTACCAGCTGACACTTTGTGTGTGTCCCATTTT
CACAATGACTGCCTTGGCCTAGGCCTTCAGTAGGGCCCATTCAATCTGAAGAT
CCATGCCCTTCAGTTCTGGACATTTTCTTTTTTGTTTTCTTTTCTTCCTTTTTAT
TTCTATTTTTAATAATTTTCTGTGTTCTCTCTCTGAGAACTCCTATTAGTCAAA
CATTTGAATCAGAATCATGGACAGTTCCTCTAATTTTCTCGTTTTTTTCTCTCT
TATTTTTCCTCTCCGTGTCTTTTAGGTCTACTATGTGAGATATTTTTTCACTTAC
TTTTCCAATAATTCTCAAAGGAGTGCTCTGATCCAATTTTGAGAATAGTCTGT
AGGGGAACGAGGGCAGGGAAGAGGCTATTGCAGTAACCCAGGTGAGAGATC
ACAGTGCTGGGGCCGGGGTAGTGCGGAGTTGGTGAAACATGCTGGGCTCCCT
GACTCCCTACATCCCCTCACATGTATGCATCCATAGCAGCCACGCCAGGGGA
AGGGTGCAACTGCCTCCTTCAACATGTGTTACAACTAGTTGTGTTCAAGGATG
TCTGATGTCTTTGTTTGCTCCATGCTACAAAGCTGCCAGCACTGTGCCACTTT
ACAGAATCCTGCTATCATTTTCTGTTTTTTGGCACTGAGATAGCACTGTCTCTC
AAACTTCTGTGACACTTGTATTTATAATTATAAAATTTGTTGCTATTCATTATT
TCTCAGTTGTTGGAAAATATTCTTTCTTTTCCAAAAGGACTGTGCTCTTTCTCA
AAAATAAAATTCCCAGTACAGTAGGTTTATTAATTTGTTTGGTCCAGTTTGAT
ATATTTTCAGACAGATTTTTGTTGTTGTTGTTGTTGCTGTTGAGATGGAATCTC
GCTCTTTTGCCCAGGCTGGAGTACAGTGGTGCGATCTTGGCTCACTGTAACCT
CTGCCTCCTGCGTTCAAGGGATTCTTCTGCCTCAGCTTCCCGAGTAGCTGGGA
CTACAGGCACGCACCACCACGCCCGGCTAATTTTTGTAATTTTAATAGAGATG
GGGTTTCACCATATTGGCCAGGCTGGTCTCGAACTCCTGACCTCAGGTGATCT
GCCCACCTTGGTCTCCCAAATTGTAGGGATTACAGGCGTGAGCCACCACCAT
GCCCAGTCTTAAGACAGATTTTCATCTTTTCCCAGCAGTTATTCAGTTGTTCA
GATCTGGAATACACCTAACCAGTCTCCCTGTACTATTTGCCCTTTGGTCCTTAT
TTAGCTCATTTTTTAAAAGAAAGTGGAGTATCAGAGTAGCTCCTTTGAATACT
CCTTTTTTTCTTTTAAATGCTTTGATTCTGAAAAACCATATACAGCTATATCTT

TTGTTTCAAGAAGTAACACTCTACCTCAGTGACTGGGATCCAGTAATAGAAA
ACACTTCAACTTCATTAACTTCACAAATAACTTATTCCTACTTCACAGAAGCA
GTGAGTACCTTGAAAACTATGAGAAGCACACAATTTTGATGCTCCCTGGGAA
ACAGATTTTAAAACCTATCGTAGTATACTAAAACTCCCATAAGAGTTCAGGC
CTTACAGACTTGGTTGGCGTTTCAGCACTCCCCTGATTTTTACTATGCAGAAG
GCTATGTTTTCATCAGGAAAATGGGGGCAAATAGTTTCCTAAGTAGATCACA
AACTGTGGGCACAGAGATTGTTTTGTTTCTTTGTTTCTCTAGCACCTAGCAGC
ATCTCCAGCACACAGTAGGTACTCAATAACATTGAACGAATTCATTTAAAATT
GATTCTATCTCCAGAACAGCAGAGGTTCCCATAATTAAAAGTCTAGTATTTGT
ACTAAAGTAGTGGTTCTTAAACTTTAGAGGGCATAAGGATCCTCTTGGAAATT
TTATAAAAATCAGGCTTCAGGGGTCCTACCTGCAGAGGTCCTAATTTGGTCAG
TCCAGGCTAGAGTCTGGGAATCTGCATTTTAAGTACATTCCTTGAATGATTTG
GGGAAGGGTAGTTTGAGGACCACTCTTTGAGATACACAATTTTTAAAAGCAT
CCTCTTTGATCCACAAAAAATACCAAAGCAAAATAGAATTTTTTTTTTTTTTT
TTGTAAAGAAAACCTTAGGGAAGAGGATTTGGATCAAACGTCAGTCAGCATA
CTAATTTTCACTTAAGTAATTTATTCAGCAGTTCCGAATTGCCTGCATTTCTTC
ATAGACATGTATATTTGTAGCCAACAAAGTGGGGAAAAGCAGCTCCACTGTC
TGAAGCGGGGCAGATGGTTTGATATTTTACTGATGGCATGGAGGGTGCTTTTA
AACCAGTTTTCCTACCAGCATTGGGCCAGATGACTGTTTCTCTAGTTGAGTAC
CAGATGAAGCAGTTGGCTTGCGTTTAAGCTCTATCTCACACACATATATATGA
TACATATATATATATATATATATATATATATATATATATAGAGAGAGAGAGAGA
GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGACT
CATATTTATTATAGTGAGAGGCTTTCAAGACCTGGGGCTAATTAAGGAAAGG
GGAATTGCGGGCTAAGTGATCAGTGCTTTTAAGTTTCCCTTTCTCTTTGTGTTT
TTTGAATGTATGGAGTTGAACGTGAACAAGTTAAATGCCTGTATAATGGAAT
GTCTCTGTGTAGTTACTGGTGTCCTTTTTAACAACGTAAGTCATGTACATTTTT
TTTTCCAGGAGGAAGAAAACAGAACTTAGACCCCAGATATATACTAACCAAA
GAGGACTTTTTATCCAATTTTGATATTTTTAAGAAAGTGAGTTGGCTTTCATTT
ACCTTGAGTTAGGAAACTGGGCTTTATTACCTGGATAGAACACTAATTAGTTC
TCAACCTCTTTCTTGAAAAGTGATGGACAAGGAAGGTATAGACCCTATATAA
TCTGATGATCTATATATGTTGGATAGCTCTCTATCCTATGATGATCTATATCTG
TTTTATAGCTCCATATCCTCTGATACGCTATACCTATTGCTTATCTGTGTATCC
TCTGATGAGCTATATCTGTTGCATATCCTCTTTCCTTTGATGATCTTTATCTGC
TGTATATCTCTCTATCCTCTGATAACCTATATCTGTTGTATATCCCTCTATTCT
CTGATGAGCTACACCTGTTGTATATCTATCTCTCTCTATCCTCTGATAACCTAT
ATCTGTTGTATATCCCTCTATTCTCTGATGAGCTATACCTGTTGTATATCTCTC
TATATCTGTCTATCCTCTGATAATCTATATCTATTGTATCTCTCTCTATATCTG
TCTATCCTCTGATAACCTGTATCTGTTGTATATCTCTCTATATCTGTCTATCCT
CTGATAACTTATATCTGTTATATCTCTCTATATATCTGTCTATCCTCTGATATT
ATAAATCTGTTGTATATCTATTGTCTGATGGACTATTTTGTATCTATCTTCTGA
TAACCTGTATCTGTTGAATATCTCGGTATCCTCTGAGGAAGTATACCTGTTAT
ATATTTCTCTCCCTCAGTAGATTAGAAAGCTGATGCAGAGAAATAAAAATAG
TAGAAACAATTATTTAGAATTACATGAATGAAGAGCTTCTTTTTCTCCCCCTA
ATCACCATGTTAACATTTTCTTTTAGAAGAAGCATGTTTTGAGCTATAAATCA
CATCCAGTGGAAGATGGTGGAACTGATAATCCTGCTTTCAACCACATTGAATT
GAACTCAGATCAGAGTGGCAAGAGCAATGGGACTCGTTTGTGAAGCTGCTCT

GATACTAGATATCCTTAAATGATGTTTCAATTTTATATGTTTTCTAAGATAATT
GGATCAGGTTTTCTTTGTGTGTGTGTGTGTTGTATCATGAGTGTTTGGGGG
ATAAGTTTTTGTTAAAACAAAGTCTGGACTATCTTCATTTACTACATCATTAA
TTGATGTTACTCTGGAGTTTAGAATTCTGGCATTGACATTTCCCTCTCTTTCCT
TTATTTCGATGAAGCTATAATTGTGAAAATTGTAACTACATAGATGCTGAAAG
GCTAATACACACATATGCACATGTATTTGATTGTCAAAGGTATATTCTTAAAT
TTGGGTATTATTGAAAATATTTTCCATGCCTTGGTGCTAGCATATAAGTTTGG
AAGTTTGCCAACATCACAATTCATCTTGAAAAGAGCTTTTTTCCCTCCTACCA
CATACACCATTCTTAGGGAGCAATGAGGTAACAGGTCTGTGTTGTCTAGATCT
TTGCTTTTTATCCCCCTATCAGTCCAGGGCATATACTAACCTGCAAACTGATT
CTGAATCAGGAAGGTGGTAATCAATAAGTATTCTGGCTGGGAAAGACCGTGG
GCCCAATGATCAAAGTCTTCTTGGTGCTGTTCATTAATTCTTGTGCCTTTTGGC
TTGTTTTCTAGAGTTTCTGGGCTTTGGCTGCTGATACTGCCTTTCTTAGACTGT
AATTTTTATCTGCATGCCCAGTTTCTGACCTATCAACTTGGGTTTTATTGTGCA
CTCTAACTGAGCTTGTCTTCATAATTTTCTGTTTATTGCCCTGGGCTTGGATAT
GTCTCAAGACACTCATGTGAATCATGCCACCCCAAATCCTGGCTTATCAAGTC
CCAGACTATAAATTATGAACTCCCATTAGCTTGGTACTAACATATACTTGATG
TAGGTATTTATGGACTTGATGATCCAAGAATATTATATTCTTCAAAATGGTTA
AGCTCCATGGAGTTAGATGACTACACTTAATGCTATTAAGTTGAACTTTTGAA
TGTCAACTAATTTGCAATCAATTAAAGATACATATGCCTAGAAATTTTGAAAT
TTCGGTATATTTATCCAGTTAAAGGGCTAAATTATATAAGCAAACACTACTTT
TTTTAAAACGTCTGGACTCAAAAAATGCTTTGTTCCATGTTTTAAAATTTTTA
AGTAGCAGTCTCAAAGTTGCTTAGCTGTTTATTTTGCTATGTTCCTAGCTAAG
AGTTTGGTTATAGGAGTTCATCAATAACTTATTTTTTGTACAGTTCCCACATTA
GATACTGTTTAAAAGTTCTTTTTTAAACTCAATTTTTTTTAGAAACATAAGAG
AAATATTTAGATACATACAAATGTTTTTATGATTAAATAATTTTATGCTTATTT
TCTGATACGTGTTATTTAGGTAATCATGCCCTGTACATTTAGAGGTTGCTAAC
TGACAATGTTAAGAAATTTTAAAAAAAAAAAAAAAGCCTGGGCATGATGGCTC
ATGCTTGTAATCTGAACATTTGGGAGGCTGAGGCAGGAAGATCGCTTGAGGT
CCAGAGTTTAAGTCCAGCCTGGAAACATAGTTAGACCTCATCTCTACAAAAA
TAAAAAATAAAAATAAAAAAAAACTTAGCTAGGCATGTTGCCACATACTTGTA
GTCCCAGTTATTAGGGAAGCTGAGGTGGGAGGATAGCTTAAGCCCAGGATTT
CAAGGCTGCATTGAGCTATGATTACACCACTGCACTCCAGCCTGGGTAACAG
AGTGAAATCTTGTCTCTGGGAAAAAAAAAAAAAAAAAAAAAAAGAGAGAGA
GAGAGAGGGAGATTTATATAACATTTAAATTAACTGCATAAACCTGGGCAAT
GTTCAAAACTCCATCTCTACAAAAAAACACAAGAATTAGCCAGGCACGGTGG
TGTGTGTCTGTAGTCTCAGCTACTTAGGAGGCTGAGGTGGGAAGATTGCTAG
AGCCAGGAGGTCGAGGCTGCACTGAGCTGTGATTGCGCTACTGTACTCCACC
CTGGGTGATGAAGCCCTAACTCAATAAATAAATAAATAAATAAATAAATAATACAA
ATAAATTAACTATATATTCATGTATTTCTTATGTGGATATATGTTATTTTTTTT
TTCCTGCTTTTCTTTTTTTTTATTTTTTTATTATACTTTAAGTTTTAGGGTACAT
GCGCACATTGTGCAGGTTAGTTACATATGTATACATGTGACATGCTGGTGCGC
TGCACCCACTAACTCGTCATCTAGCATTAGGTATATCTCCCAATGCTATCCCT
ACCCCCTCCCCCCACCCTACCACAGTCCCCAGAGTGGATATTCCCCTTCCTGT
GTCCATGTGATCTCATTGTTCAATTCCCACCTATGAGTGAGAATATGCGGTGT
TTGGTTTTTTGTTCTTGCCATAGTTTACTGAGAATGATGATTTCCAATTTCATC

CATGTCCCTACAAAGGACATGAACTCATCATTTTTTATGGCTGCATAGTATTC
CATGGTGTATAAGTGCCACATTTTCTTAATCCAGTCTATCATTGTTGGACATTT
GGGTTGGTTCCAAGTCTTTGCTATTGTGAATAATGCTGCAATAAACATACGTG
TGCATGTGTCTTTATAGCAGCATGATTTATAGTGCTTTGGGTATATACCCAGT
AATGGGATGGCTGGGTCAAATGGTATTTCTAGTTCTAGATCCCTGAGGAATC
GCCACACTGACTTCCACAGTGGTTGAACTAGTTTACAGTCCCACCAACAGTGT
AAAAGTGTTCCTGTTTCTCCACATCCTCCCCAGCACCTGTTGTTTCCTGACTTT
TTAATCATTGCCATTCTAACTGGTGTGAGATGGTATCTCATTGTGGTTTTGATT
TGCATTTCTCTGATGGCCAGTGATGATGAGCATTTTTTCATGTGTTTTTTGGCT
GCATAAATGTCTTCTTTTGAGAAGTGTCTGTTCATGTCCTTCGCCCACTTTTTG
ATGGGGTTGTTTGTTTTTTTCTTGTAAATTTGTTTGAGTTCATTGTAGATTCTG
GATATTAGCCCTTTGTCAGATGAGTAGGTTGCAAAAATTTTCTCCCATTTTGC
AGGTTGCCTGTTCACTCTGATGGTAGTTTCTTTTGCTGTGCAGAAGCTCTTTAG
TTTAATTAGATCCCATTTGTCAATTTTGTCTTTTGTTGCCATTGCTTTTGGTGTT
TTGGACATGAAGTCCTTGCCCATGCCTATGTCCTGAATGGTAATGCCTAGGTT
TTCTTCTAGAGTTTTTATGGTTTTAGGTCTAACGTTTAAGTCTTTAATCCATCT
TGAATTGATTTTTGTATAAGGTGTAAGGAAGGGATCCAGTTTCAGCTTTCTAC
ATATGGCTAGCCAGTTTTCCCAGCACCATTTATTAAATAGGGAATCCTTTCCC
CATTGCTTGTTTTTCTCAGGTTTGTCAAAGATCAGATAGTTGTAGATATGAGG
CGTTATTTCTGAGGGCTCTGTTCTGTTCCATTGATCTATATCTCTGTTTTGGTA
CCAGTACCATGCTGTTTTGGTTACTGTAGCCTTGTAGTATAGTTTGAAGTGAG
GTAGCGTGATGCCTCCAGCTTTGTTCTTTTGGCTTAAGATTGCCTTGGCAATG
CGGGCTCTTTTTTGGTTCCATATGAACTTTAAAGTAGTTTTTTCCAATTCTGTG
AAGAAAGTCATTGGTAGCTTTATGGGGATAGCATTGAATCTGTAAATTACCTT
GGGCAGTATGGCCATTTTCACGATATTGATTCTTCTTACCCATGAGCATGGAA
TGTTCTTCCATTTGTTTGTATCCTCTTTTATTTCCTTGAGCAGTGGTTTGTAGTT
CTCCTTGAAGAGGTCCTTCACATCCCTTGTAAGTTGGATTCCTAGGTATTTTAT
TCTCTTTGAAGCAATTGTGAATGGGAGTTCACTCATGATTTGGCTCTCTGTTT
GTCTGTTGTTGGTGTATAAGAATGCTTGTGATTTTGGTACATTCATTTTGTATC
CTGAGACTCTGCTGAAGTTGCTTATCAGCTTAAGGAGATTTTGGGCTGAGTCA
ATGGGGTTTTCTAGATATACAATCATGTCGTCTGCAAACAGGGACAATTTGAC
TTCCTCTTTTCCTAATTGAATACCCTTTATTTCCTTCTCCTGCCTGATTGCCCTG
GCCAGAACTTCCAACACTATGTTGAATAGGAGCGGTGAGAGAGGGCATCCCT
GTCTTGTGCCAGTTTTCAAAGGGAATGCTTCCAGTTTTTGCCTATTCAGTATG
ATATTGGCTGTAGGTCTGTCATAGATAGCTCTTATTATTTTGAAATACATCCC
ATCAATACCTAATTTATTGAGAGTTTTTAGCATGAAGGGTTGTTGAATTTTGT
CAAAGGCTTTTTCTGCATCTATTGAGATAATCATGTGGTTTTTGTCTTTGGCTC
TGTTTATATACTGGATTACATTTATTGATTTGCATATATTGAACCAGCCTTGCA
TCCCAGGGATGAAGCCCACTTGATCATGGTGGATAAGCTTTTTGATGTGCTGC
TGGATTCGTTTTGCCAGTATTTTATTGAGGATTTTTGCATCAATGTTCATCAAG
GATATTGGTCTAAAATTCTCTCTTTTGGTTGTGTCTCTGCCCGGCTTTGTTATC
AGAATGATGCTGGCCTCATAAAATGAGTTAGGGAGGATTCCCTCTTTTTCTAT
TGATTGGAATAGTTTCAGAAGGAATGGTACCAGTTCCTCCTTGTACCTCTGGT
AGAATTCGGCTGTGAATCCATCTGGTCCTGGACTCTTTTTGGTTGGTAAACTA
TTGATAATTGCCACAATTTCAGCTCCTGTTATTGGTCTATTCAGAGATTCAAC
TTCTTCCTGGTTTAGTCTTGGGAGAGTGTATGTGTCGAGGAATTTTTCCATTTC

TTCTAGATTTTCTAGTTTATTTGCGTAGAGTTGTTTGTAGTATTCTCTGATGGT
AGTTTGTATTTCTGTGGGATCGGTGGTGATATCCCCTTTATCATTTTTTATTGT
GTCTATTTGATTCTTCTCTCTTTTTTTCTTTATTAGTCTTGCTAGCGGTCTATCA
ATTTTGTTGATCCTTTCAAAAAACCAGCTCCTGGATTCATTAATTTTTGGAAG
GGTTTTTTGTGTCTCTATTTCCTTCAGTTCTGCTCTGATTTTAGTTATTTCTTGC
CTTCTGCTAGCTTTTGAATGTGTTTGCTCTTGCTTTTCTAGTTTTTTTTTTTTTA
TTATTATACTCTAAGTTTTAGGGTACATGTGCACATTGTGCAGGTTAGTTACA
TATGTATACATGTGACATGCTGGTGCGCTGCACCCACCAACGTGTCATCTAGC
ATTAGGTATATCTCCCAATGCTATCCCTCCCCCCTCCCCCGACCCCACCACAG
TCCCCAGAGTGTGATATTCCCCTTCCTGTGTCCATGTGATCTCATTGTTCAATT
CCCACCTATGAGTGAGAATATGTGGTGTTTGGTTTTTTGTTCTTGCGATAGTTT
ACTGAGAATGATGGTTTCCAATTTCATCCATGTCCCTACAAAGGACATGAACT
CATCATTTTTTATGGCTGTATAGTATTCCATGGTGTATATGTGCCACATTTTCT
TAATCCAGTCTATCATTGTTGGACATTTGGGTTGGTTCCAAGTCTTTGCTATTG
TGAATAGTGCCGCAATAAACATACGTGTGCATGTGTCTTTATAGCAGCATGAT
TTATAGTCCTTTGGGTATATACCCAGTAATGGGATGGCTGGGTCAAATGGTAT
TTCTAGTTCTAGATCCCTGAGGAATCGCCACACTGACTTCCACAATGGTTGAA
CTAGTTTACAGTCCCACCAACAGTGTAAAAGTGTTCCTATTTCTCCACATCCT
CTCCAGCACCTGTTGTTTCCTGACTTTTTAATGATTGCCATTCTAACTGGTGTG
AGATGATATCTCATAGTGGTTTTGATTTGCATTTCTCTGATGGCCAGTGATGA
TGAGCATTTCTTCATGTGTTTTTTGGCTGCATAAATGTCTTCTTTTGAGAAGTG
TCTGTTCATGTCCTTCGCCCACTTTTTGATGGGGTTGTTTGTTTTTTTCTTGTAA
ATTTGTTTGAGTTCATTGTAGATTCTGGATATTAGCCCTTTGTCAGATGAGTA
GGTTGCGAAAATTTTCTCCCATGTTGTAGGTTGCCTGTTCACTCTGATGGTAG
TTTCTTTTGCTGTGCAGAAGCTCTTTAGTTTAATTAGATCCCATTTGTCAATTT
TGGCTTTTGTTGCCATTGCTTTTGGTGTTTTGGACATGAAGTCCTTGCCCACGC
CTATGTCCTGAATGGTAATGCCTAGGTTTTCTTCTAGGGTTTTTATGGTTTTAG
GTCTAACGTTTAAGTCTTTAATCCATCTTGAATTGATTTTTGTATAAGGTGTAA
GGAAGGGATCCAGTTTCAGCTTTCTACATATGGCTAGCCAGTTTTCCCAGCAC
CATTTATTAAATAGGGAATCCTTTCCCCATTGCTTGTTTTTCTCAGGTTTGTCA
AAGATCAGATAGTTGTAGATATGCGGCATTATTTCTGAGGGCTCTGTTCTGTT
CCATTGATCTATATCTCTGTTTTGGTACCAGTACCATGCTGTTTTGGTTACTGT
AGCCTTGTAGTATAGTTTGAAGTCAGGTAGTGTGATGCCTCCAGCTTTGTTCT
TTTGGCTTAGGATTGACTTGGCGATGTGGGCTCTTTTTTGGTTCCATATGAACT
TTAAAGTAGTTTTTTCTAATTCTGTGAAGAAAGTCATTGGTAGCTTGATGGGG
ATGGCATTGAATCTGTAAATTACCTTGGGCAGTATGGCCATTTTCACGATATT
GATTCTTCCTACCCATAAGCATGGAATGTTCTTCCATTTGTTTGTGTCCTCTTT
TATTTCCTTGAGCAGTGGTTTGTAGTTCTCCTTGAAGAGGTCCTTCACATCCCT
TGTAAGTTGGATTCCTAGGTATTTTATTCTCTTTGAAGCAATTGTGAATGGGA
GTTCACTCATGATTTGGCTCTCTGTTTGTCTGTTGTTGGTGTATAAGAATGCTT
GTGATTTTGGTACATTGATTTTGTATCCTGAGACTTTGCTGAAGTTGCTTATCA
GCTTAAGGAGATTTTGGGCTGAGTCAATGGGGTTTTCTAGATATACAATCATG
TCGTCTGCAAACAGGGACAATTTGACTTCCTCTTTTCCTAATTGAATACCCTTT
ATTTCCTTCTCCTGCCTGATTGCCCTGGCCAGAACTTCCAACACTATGTTGAA
TAGGAGCGGTGAGAGAGGGCATCCCTGTGTTGTGCCAGTTTTCAAAGGGAAT
GCTTCCAGTTTTTGCCCATTCAGTATGATATTGGCTGTGGGTTTGTCATAGAT

AGCTCTTATTATTTTGAAATACGTCCCATCAATACCTAATTTATTGAGAGTTTT
TAGCATGAAGGGTTGTTGAATTTTGTCAAAGGCTTTTTCTGCATCTATTGAGA
TAATCATGTGGTTTTTGTCTTTGGCTCTGTTTATATGCTGGATTACATTTATTG
ATTTGCGTATATTGAACCAGCCTTGCATCCCAGGGATGAAGCCCACTTGATCA
TGGTGGATAAGCTTTTTGATGTGCTGCTGGATTCGGTTTGCCAGTATTTTATTG
AGGATTTTTGCATCAATGTTCATCAATGATATTGGTCTAAAATTCTCTTTTTTG
GTTGTGTCTCTGCCTGGCTTTGGTATCAGAATGATGCTGGCCTCATAAAATGA
GTTAGGGAGGATTCCCTCTTTTTCTATTGATTGGAATAGTTTCAGAAGGAATG
GTACCAGTTCCTCCTTGTACCTCTGGTAGAATTCGGCTGTGAATCCATCTGGT
CCTGGACTCTTTTTGGTTGGTAAACTATTGATTATTGCCACAATTTCAGCTCCT
GTTATTGGTCTATTCAGAGATTCAACTTCTTCCTGGTTTAGTCTTGGGAGAGT
GTATGTGTCGAGGAATGTATCCAATTCTTCTAGATTTTCTAGTTTATTTGCGTA
GAGTTGTTTGTAGTATTCTCTGATGGTAGTTTGTATTTCTGTGGGATCGGTGGT
GATATCCCCTTTATCATTTTTTATTGTGTCTATTTGATTCTTCTCTCTTTTTTC
TTTATTAGTCTTGCTAGCGGTCTATCAATTTTGTTGATCCTTTCAAAAAACCAG
CTCCTGGATTCATTGATTTTTGGAAGGGTTTTTTGTGTCTCTATTTCCTTCAGT
TCTGCTCTGATTTTAGTTATTTCTTGCCTTCTGCTAGCTTTTGAATGTGTTTGCT
CTTGCTTTTCTAGTTTTTTTAATTGTGCTGTTAGGGTGTCAATTTGGATCTTTC
CTGCTTTCTCTTGTGGGCATTTAGTGCTGTAAATTTCCCTCTACACACTGCTTT
GAATGCGTCCCAGAGATTCTGGTATGTTGTGTCTTTGTTCTCGTTGGTTTCAA
AGAACATCTTTATTTCTGCCTTCATTTCGTTATGTATCCAGTAGTCATTCAGGA
GCAGGTTGTTCAGTTTCCATGTAGTTGAGCGGTTTTGAGTGAGATTCTTAATC
CTGAGTTCTAGTTTGATTGCACTGTGGTCTGAGAGATAGTTTGTTATAATCTC
TGTTCTTTTACATTTGCTGAGGAGAGCTTTACTTCCAAGTATGTGGTCAATTTT
GGAATAGGTGTGGTGTGATGCTGAAAAAAAATGTATATTCTGTTGATTTGGGG
TGGAGAGTTCTGTAGATGTCTATTAGGTCCGCTTGGTGCAGAGTTGAGTTCAA
TTCCTGGGTATCCTTGTTCACTTCCTGTCTCGTTGATCTGTCTAATGTTGACAG
TGGGGTGTTAAAGTCTCCCATTATTAATGTGTGGGAGTCTAAGTCTTTTTGTA
GGTCACTCAGGACTTGCTTTATGAATCTGGGTGCTCCTGTATTGGGTGTATAT
ATATTTAGGATAGTTAGCTCTTCTTGTTGAATTGATCCCTTTACCATTATGTAA
CGGCCTTCTTTGTCTCTTTTGATCTTTGTTGGTTTAAAGTCTGTTTTATCCGAG
ACTAGGATTGCAACCCCTGCCTTTTTTTGTTTTCCATTTGCTTGGTAGATCTTC
CTTCATCCTTTTATTTTGAGCCTATGTGTGTCTCTGCACGTGAGATGGGTTTCC
TGAATACAGCACACTGATGGGTCTTGACTCTTTATCCAATTTGCCAGCCTGTG
TCTTTTAATTGGAGCATTTAATCCATTTACATTTAAAGTTAATATTGTTATGTG
TGAATTTGATCCTGTCATTATGATGTTAGCTGGTGATTTTGCTCGTTAGTTAAT
GCAGTTTCTTCCTAGTCTCGATGGTCTTTACATGTTGGCATGATTTTGCAGCG
GCTGGTACCGGTTGTTCCTTTCCATGTTTAGCGCTTCCTTCAGGAGCTCTTTTA
GGGCAGGCCTGGTGGTGACAAAATCTCTCAGCATTTGCTTGTCTGTAAAGTAT
TTTATTTCTTCTTCACTTATGAAGCTTAGTTTGGCTGGATATGAAATTCTGGGT
TGAAAATTCTTTTCTTTAAGAATGTTGAATATTGGCCCCCACTCTCTTCTGGCT
TATAGGGTTTCTGCCGAGAGATCTGCTGTTAGTCTGATGGGCTTCCCTTTGAG
GGTAACCTGACCTTTCTCTCTGGCTGCCCTTAACATTTTTTCCTTCATTTCAAC
TTTGGTGAATCTGACAATTATGTGTCTTGGAGTTGCTCTTCTTGAGGAGTATC
TTTGTGGCGTTCTCTGTATTTCCTGAATCTGAACGTTGGCCTGCCTTGCTAGAT
TGGGGAAGTTCTCCTGGATAATATCCTGCAGAGTGTTTTCCAACTTGGTTCCA

TTCTCCCCATCACTTTCCGGTACACCGATCAGACGTAGATTTGGTCTTTTCAC
ATAGTCCCATATTTCTTGGAGGCTTTGCTCATTTCTTTTTATTCTTTTTTCTCTA
AACTTCCCTTCTCGCTTCATTTCATTCATTTCATCTTCCATCGCTGATACCCTT
TCTTCCAGTTGATCGCATTGGCTCCTGAGGCTTCTGCATTCTTCACATAGTTCT
CGAGCCTTGGTTTTCAGCTCCATCAGCTCCTTTAAGCACTTCTCTGTATTGGTT
ATTCTAGTTATACATTCTTCTAAATTTTTTTCAAAGTTTTCAACTTCTTTGCCTT
TGGTTTGAATGTCCTCCCATAGCTCAGAGTAATTTGATCGTCTGAAGCCTTCT
TCTCTCAGCTCGTCAAAGTCATTCTCCATCCAGCTTTGTTCCGTTGCTGGTGAG
GAACTGCGTTCCTTTGGAGGAGGAGAGGCGCTCTGCGTTTTAAAGTTTCCAGT
TTTTCTGTTCTGTTTTTTCCCCATCTTTGTGGTTTTATCTACTTTTGGTCTTTGA
TGATGGTGATGTACAGATGGGTTTTTGGTGTGGATGTCCTTTCTGTTTGTTAGT
TTTCCTTCTAACAGACAGGACCCTCAGCTGCAGGTCTGTTGGAATACCCTGCC
CTGTGAGGTGTCAGTGTGCCCCTGCTGGGGGGTGCCTCCCAGTTAGGCTGCTC
AGGGGTCAGGGACCCACTTGAGGAGGCAGTCTGCCCGTTCTCAGATCTCCAG
CTGCGTGCTGGGAGAACCACTGCTCTCTACAAAGCTGTCAGACAGGGACATT
TAAGTCTGCAGAGGTTACTGCTGTCTTTTTGTTTGTCTGTGCCCTGCCCCCAGA
GGTGGAGCCTACAGAGGCAGGCAGGCCTCCTTGAGCTGTGGTGGGCTCCACC
CAGTTCGAGCTTCCCGGCTGTTTTGTTTACCTAATCAAGCCTGGGCAATGGCG
GGCGCCCCTCCCCCAGCCTCGCTGCCGCCTTGCAGTTTGATCTCAGACTGCTG
TGCTAGCAATCAGCGAGATTCCGTGGGCGTAGGACCCTCCGAGCCAGGTGCA
GGATATAATCTCGTGGTGCGCCGTTTTTTAAGCCGGTCCGAAAAGCGCAATAT
TCGGGTGGGAGTGACCTGATTATCCAGGTGCGTCTGTCACCCCTTTTTTTGAC
TCGGAAAGGGAACTCCCTGTCCCCTTGCGCTTCCCAAGTGAGACAATGCCTC
GCCCTGCTTCGGCTTGCGCATGGTGCACGCACCCACTGACCCGCGCCCACTGT
CTGGCACTCCCTAGTGAGATGAACCCTGTACCTCAGATGGAAATGCAGAAAT
CACCTGTCTTCTGCGTCGCTCACGCTGGGAGCTGTATACCGGAGCTGTTCCTA
TTTGGCCATCTTGGCTCCTCCCCCGGTATTCCTTCTTTTCTTCCACTGTGAGAG
TTACTTAAAGCTCGGCGTCCGTGATGGTCTAGGGGGCTTCTGAGGCGATCGG
GCAGTGTCCGTCTTCAGCCGCTAAGCCGAGAAGATCTGGGAAGGAGTCAGTC
AGAGAGCCTTGGGCCAGAGTTCCAGGGCCTCTGGGAGTGGCTGCCAGGTGAG
TTGAACAGTCCGATTTTCAGTGGGGTCCCACACAGATGGGACATGGCTTAGG
AGGAATCCCAGGCTGTGGGCATTCCTTGGCCCAGTGGCCAGATTCGATATAT
GTTATTTTTAAATCACTGTATTTGTAAGCAAATATCAAATTTAGGGAAGTCTT
TCTACAATGTTTTAATAAGTAGAAAGATATGTTTGTTTTACATGAATGTGTTT
TGAACTATGGTTATTTGTTTAATAATTCTAAATGCATATGTGTGTAAAATGCT
TCAATTTTGGAAATCAAAGTCAGGCCATTTTTTTGTCTTACCTGATTGCCAGG
GAGTTACGCCATGTATTCTTAATGAGAAACATGATGTTTCCATTCTTGTTCAC
TTTCCTTTAGACAGAATATATTTTTGTGACATTTAGAACTATCAATATTTTAGT
TTTATAAACACAGGAGAATGCCTGATAGAATTCTTAAGAAAGCAATGTAACA
GTATTAGCTCAAAATAATTTATCTTAATTTCTAAATTTTTAGATAAAACCAAA
TAAGGGTTAAATGTTAATCCATTGTCACTTAAATTACATAATCTGCTACTCTT
AGTTATTTGAATGACAAAAACACCAGTGGGGGAAAAACCATACAAGTTGTCA
ATGTCTGTTTTGCTGTTGACAAGTTGTATACCCTTAAATTGACCCCTAATCTCC
TCTAACAATGGTACATAGCACTAAGCTCCTACCTACCTCACAGAAATAATGTC
AGTAGAAAAGATCAGGTTGAGTTCTTTGGCAGAATAGCACTTTACTAACTCA
AATAGTGTTACTTAATATTTCAATATGATTGGGAATCAAAGTTTGAGACAAA

AGTCATTTGCCAGTTTTAAAAAATAGAGCTGTTAATTTGCAATATCATGATGT
AGAGATAGTGCCTTCTCTTAAAAATGTGTGTCATGGAAATAGTAAAATATATT
TAGGAGTCAGCAGGATTATTCCAACAGAGGGAGTGTAAACTTTAAAGAAAAA
TATGATTCGGGAGGCTGAGGTGGGTGGATCATGAGGTCAGGAGTTCGAGACC
AGCCTGGCCAACATAGTGAAACCCCGTCTCTACTAAAAATACAAAAATTAGC
TGGGCATGGTGGCACACACCTGTAGTCCCAGCTACTCGGGAGGCTGAGGCAG
GAGAATCGCTTGAACCTGGGAGGTGGAGGTTGTGGTGAGCCGAGATCACACC
ACTGCACTCCAGCCTGGGCAACAGAGCGAGATTCCATCTCAAAAAATACATA
TATATTTTTGACATATATAATATATATATGTCAGTAATATTCACCCCATAGAA
AATGAAAATTTTATAGGAAAGATGTAAAACAGCATAAATTCACATTCATCTT
ATTAGTTGCTTATGCAATCATTTTCTCTCCAGATCATTGGTTCTCAAAGGGGA
CAATTTTGCCTCTCAGGGGATATTTGGAAATGTCTGGAGACATTTTTGGATGG
CACTAGTGGTATCTAGTAGGTAGAATTTAGGGAAACTGGTAAACATCTCCTG
AGGCCTGCAAGGGTGCTCTCCTCCTCCACAACAAAAAATTATCCAGCCTAAG
ATGTCCATAGTGTAGAGTTGGAGAAACCTTGCCCTGGAGAATAAGGTTGATT
TTCTTGAAGTCACACAGCCTGGTTGTGTTTCTATAGGGAAACAGCCTGAAAAT
TCTATCTGAATGTTCTCATCTACAGGTAAGGATGAAAATGCCACTGGCATATC
TAATATTATGATGCAGAACAATGACCATGTATTTTCACAGCATTATGAAATTA
TTAAGGACCATAGAATTGTGAATAATTATTTAAAGAAGTCTTAGGACAGTTT
AGATTCTCCACATGCCTTCTAATATTGACACACATTAGGATGAAGGAAATATT
AAATACATACATGTAAAGATTTTGAATTTTTTTTCAACTGAGCGTCCAGGATA
TAAATACAAGGAACAGGGAGGGGGTTGAGATGGCGGAAGTAACTCTGTATT
GATTCTTATAGGAAATTCTGAGTTTTTCCATAAAGACAAAGAGTTTATTGAGT
ACATGAGCATTTAGTTACTGAAAATTCACTGTATGCTTTTCTAAGTTTTGAGC
TTATTGTTTATGAAATCCTTGAGAAAGTTGAACATTTCAATGTAAAAACATGG
TTGTGAATCTGAATTTTCAACTTGCTGATTAAACTCCCTGCAAGTTTCTTTGCA
GTTGTCTGTTTTGGGGGGATAAATGTCAAATTGAATACAGTTAATTTTATCAG
CCTTTACAAAAAGATACTTCCACCCTATTTACAACATAAAGGACTATTCCTAA
GTGCTGTCTGTAGATTACAAAAAGTATAAACATGTAGAATTTTTGTCACAGA
AGACTATTTTATTTTTAATGAATTAACACCGTATTGAAAAATAAAAAGTACAA
AAAAGTACAAACTTTTCTGTCCCAATACATTATAAAACGTTTTATTTTAATAG
CTTTAGAGGTACAGTTTTTGGTTACATGGGTGAATCGTATAGTGGTGAAGTGT
GAAGGTTCAGTGCACTGGTCACCTGAGTAGTGTACATGGTGCCCAATAGATA
GTTTTTCATTCCTCTTCCTCAGCCTCCCCACCTTCTGAGTCTCTAATGTCCATT
ATACCACTCTGTATGCCTTTGTGTACTCATAGCTTAGCTCCCACTTACAAATG
AAAACATGTGGTATTTGCTTTTCCATTCCTGACTTATGTCACTTAGAATAATA
GCCTCCGGTTCCATCTAAGTTGCTGCAATAGACATTATTTCATTCTTTTTTATC
CCTGAGTAGTACTCCATGGTGTATGTGTATGTATATACATATATATATATATA
TATCTCACATTTCATATATATACTCACATTTTTTATCCACTCATCAGTTGATAG
GCACTTAGGTTGATTCCATATCCTTGCAATTGTGAATCGTGCTGCGATAAACA
TGTGCATACAGGTGTCTTTTTGACATAGTGACTTCTTTTCCTTTAGGCAGATAC
CCAATAGTTGTTCCAATCCAATTTTTAATTGGGGTAATTTAATCTTTTAAAAG
TTGGTCCAAGTTAATTGTTGATAATATCAGGACTTTAAAAGAGAAACAGAAG
TTCTTAACCTGAGTGTTTTTTCTTTCTTTTGAAAAAATATCAGTTTGAAGTTTT
AAATTTCTATTTTATATCTCAAAGCTATAGTTTTGCTTGTGGGGTATAAAATT
AAGTGGACAACTAAGACAGAGAACTTAGGTGCCAAAGATGACCATGTTTATA

CTCAATCACCCAATTTGGAACCACATCATCAAAGAAGCAGTTGCCAGTGTTC
CCCCTAGTGTGAAGTTTCCACTTCTCTCAGTTAAAGCACCTGTCTGTCATCTC
ATTTAAAGCACCTACTTACTTCCTACCTATTCAAGTCTTGATTAAGCAAAATG
CAGATTTTCCATATACAGGAAATTTGGCATAACCTTTCACTTTAAAGGTCAAA
TCAGGTCTCCATCATTTAAATTCATCAAAGAAAGAATATTTTGAAGTTGTTGA
CTTTGTTACTCATTCCCATTTTGCAATCATGTATTGTTATTCCCTTCCTCATTTA
AAAAGGCTTCTTTTACCCCTTACCCTTGTTTAGGCTGCACCACCAAAGGTCAT
TGGATATCAATGGATGGGATTCACTCCTGGAGCTCCAGACTCACTCACACAT
GTGCATCAAGGATTCAGGATTCTCTCCATTTCTGCTTTCCTTAACTTTCCAAAG
CCAGACCTTTATTCTCTTCTGTATTAGGATCTGGTCTGTCACTGGGCTTTTCTC
ATTCTATTCTAGAGCTTCTTAAACTTCAGTGTTCATCAGAAGCACCTGGAGGA
GCTGGTTAAAACACAGATTGCTGGGCTTCACCCCAGAGTGTCTGATTTAACA
GCTCTCAGGTGGGTCCTGAGAATTTACCTTTCTCAAAATTTTCCTGAGGATGA
TGGTGCTTCTGGTCTGGGAGTCACACTTTGGAAACTACTGTTCCAGTCCACAG
TTGTCTCTTTGGAAACCGAATCTGATGATTCACTCCTCTGCTTGAAGATCTCT
ATGACTACAGAATAAAAGCCCCATCCCTTAGCCTGATGGGCTTCTCAGAAGT
ATTTATTGGTACCCTCCTTCACATGTTACATAGGCTTGTCACATTGAGCTTCTC
ATATGTGCTAAATATACCACCATTTTCTTGCCTTCTTGTTATTTTACATGCTAT
CCTCTTTGTCTAGGCTACCCATCCTCTGTCTACTTCTCAGATCTTTGAAAAACA
CCTGCTCAGTTGTTAGAACCCAGCTTACCTATCACTTCTCTAACTCTTGACAC
ATTCCATGGGTGATCGTGATCTTATACTTACCTCCGGCTCTAGTCATTTTGTTG
TACTGTACATGTATTAATGTACACAGCTATCTCTTAGGTGGCACATAGTCTCT
ATTCCTGATGTTTCCATCCAGGTGGATGAACTGTCCATTAGAGTAACTTTCTG
GATCTCTCTCTGCCCCTTTCCTGCTTATTCTCCCTATGTAAACAGGAAGTGACT
TTTGTGATCAGTAAGTCTGAGAGAGGAATCAGACATGTATATCTTAGTCCTTT
CCACTTCCATTTCTTTTTGGCATCTGCCTGCTTAAAGAATATGCATGATCTATG
CCTTACAACTCCTTGCTCCCATGATCTCTTTGACCTCTATTACTCCATGCCAGG
TCTTGCCATTCCTTAAAAACATGTTCCCACCTCACGGTCATGTGCATTGCTTA
CAACACCCTACTCAATATCCATTTGGCTCACTCTTTCAGCTCCTACAGGTCTTT
ATTCAGATGTCATCTTCTAGGTGAGGTATTCTCTGATCTCTATTTAAAATTGC
AACTTTCCTCCGCCATGCACCCTATCCCCCTTGCTTGCTTTATTTCTCTCCCAT
CTCTATTATCATTGAACACACAATATTTTACTTGTTTGTTGTATGTCATTCCCC
AATAAAATAAAAACTCCAAGAGGTGAGGATTTTTGCTGGTTCTGTTTAGTAAT
TTCTCTAGCAGATGTAGAACATGGAAGGCACTCAATACAAATTGGAATACAT
GCTTTTGGTCATGAGATAAGGGTTAGTGATAAAAATAGCCTGCTTCCATAGG
GATGCTTGGGGTCTTGACACCAGCCGGTGACTAGATATGTGTAATTCTCAGAT
TTAGTGTTAGGGAAACTTTGTTGACTTGTAGTTAGTCATGTCTTCCAATCATC
CATTACCAATAATATTAGTAATATTGTAATAAATAAGAGACTCATCTCTACCA
TCACTGAGTTTATTGTCTAATACAGAAAATGGGCAAAATACAAGTAGTTACA
GTAAAGTGTTGTAACCTGAACACAGATGTGCCTGCTCGCCACTTGAAAACTA
AAATAAAGAGAGAAGAGAGTTGGTGGGAGGAAACGCAGGTTTATTTGGAGA
ACCAGCAGACCAAGAAGATGATAAACTGTTGTCCTAAAGTACCATCTTAAGT
CAGTACAAATTGCAGATTATTTTTATGTTAAGAACAGGGGGAAGGAAAGGTG
GGTGGGATCAAGAGGTGACTGACAACTGCAGACATCTGGGCACCAACAAGG
GTCTGAGGAGGTTGAGAACTTCTATTTCCTTGGTCAGGTCACAATGCTCTTAT
AAATATTTAACAAAACATAGTTGTTTACATACTTTCCCTTTAATCACAGAGTT

AGTTTCAAAAACTACATGATTGTTTCTTTGCATATGATATGGTTTGTATTTATG
TCCCCACTCAAATTTCATGTGGAATTGTAATCCCTACTGTTGGAGAAGAGGCC
TGCTGGAAGTTGATTGGATCATGAGGCCGACTTCCCCATTGCTGTTCTTGTGA
TAATGAATGAGTTCTCATGAGATCCGGTTGTTTAGAAGTGTGTAGCACCTCCC
CTTTTGCTCTTTTGCCTCCTGCTCCAGCCATGTAAGATGTGCCTCCTTCCTCTT
TGCCTTCTGCCATGATTGTAAGTTTCCTGAGGCCTCCTCAGCCATGCTTTCTGT
ACAGCCTGCAGAATCATGAGCCAATTAAACCTCTTTGCTTTATAAATTACCCA
GTCTCAGGTAGTTTCTTACATTTAATAGCAATGCGAGAACGGACTAATTCAGC
ATATTATCTCACTGCTCTAAAATGATCCTAACCTACATGCAGGAATGGGTAAA
GGCTCCTTAAACAAAAATGGAGTTATATATGTTAGTTCTTTTGCTGTTTCACT
GTTACAGTGTGGTAAGTATTGCAATTGGAGCATTGCATGTGCTATAATCCAAA
CACGTGCTAAGTGACATAAGTATCAACGAGGGAGTAACAGGGATGGACAAA
AAGGGACCAAATCCAGCTATAGAATACTTTCCTGACTAGATGAGGAATAAGC
TCAGATTTGAAAGATGTATAGTCATTAGCTAAGCAAAGGAAAGGAAAAGAA
GAAGTGGTTTAGGCAAAGGGAACCACATGTTCTAAAGCCTAGAGGACTGAGG
GATCATGGTGCATAAGAAGAGTATGTATATAGGGTAGAGTGAATGATAACAT
GGGCACCAGCCAGACTATGGGAAGTCATGCTTGAGGTTTTAGACTTTATCCTT
ATGGGGATAAGAAACCACTGAAGAGTTGTAGGAAGAGACGTACGATATGAT
CAAATTTGGATTTCTGAAAGTTCAGTTCACCATAGCTATAAAGTGAAGAATG
GATGGGTGGAAGGGAGTATGCCTGAGGATCAGGAGACTATTTAGGAGTCTGT
GTTGTAGTCTTTGTGAGAGAAAATGGTGGCCTAGATTATGATGGTGGTAATG
AGGATGAAGAAAGATAGATACATGTGAGAGGATTGAATTGACAGGACTTGG
TACATTATTGGAAAGAAAGATGTCAGAAATTACTTATGTTTTTCTGCTTAAGC
ATTTTGGGTGGACATGGAGCCATTCAAAAGGTAGTTACCATAAAAAACTTGG
TGTAGGATAGGATGCTCAGTTTCAGACCTATTGGGTTCCAGCAGCAATTCTCA
ACCTAGGAAAGTTCTTAAAAATAAAATAAATTAAATCTGTGAGTTATTAGTG
ATACTAAAATCATAATCAATTTGTAACCATTCCATTTATCCTTTTCTTCTCCTT
CTTCTGCTCCTTCTTCATCATTGTCCTCATCTCCTTCTCCTTCTACTATTGTTTT
ATTCATGTTGGGGTTCCATAGACTATTAGTAGCTGATTATTTGTGGGAAGGGA
AGGTATTTGAGTGTCTTACTAATTTTCTTTTATTAGGTTGGTACAAAAGTAATT
GCAGTTTTTGTTAATATTATTTTTCTGACATATTTTCCAGAAACTTCAAATTTT
CTGCTAAATTTTTCAGAACATTTGTTAATCTGAAAATATTAGATTAACTCTCA
TATGACAATTTATCTGGTTAACAAATTTTAGTTTGAGAGTTCACCCAACATTT
TTAAATATTCTTCCATTATCTTTTAGTATCTATTCCTGATGTTGAGATGTTTGC
TGTCTGCCTAGTTGTTATCCCTTGGAGATACAGTAGCCTTTGATATATGTGGA
TAATTGGTTCTAGGTCTCCCAACGTAACCAAAATTTGCATATACTCAAGTCCT
GTATTCAGCCCTGCAGAACTCGTGTATACAAAAAGTTGACTCTCCATATATGC
AAGTTTCACCTCCTGCAAATACTTCATTTTTGGTATATTTTCGATCTGCATTCA
TTTGAAAAAAAAATCCATGTAAAAGTGGAACCTTCCAGTTCAAACCTGTGTTGT
TCCAAGGGGCAACTGTAATCTCTCTTTTCTTTGTGAAAGCATTTAAAATTTAC
TCCTTATTGTGCTTTTTCTCTAGTTTTGCCACAATGAGTCTAGGTATGAATTTG
TTTTTATTTCCTCCTTGGGACTTTTGTTTCTTCAATGTAGAATCATACCTTTAA
TATTGGAAATATTTTTGGTATTATGTCTGAATATTTCCCCTTTCCCATATTTTT
TCTATTCTGTACCTCTTTAATTCCTGTTAGTTGTATGTTGTACCTTTTAAATCCT
GTCCTTCGTATCTCCTAATTTGGTTCATCTTTTCCATCTTTTTGTATTTATGTGA
TGCATCCTGGACAATTTTCTCAGATATTTTCTTTCATTCCCCTTCCCAGCTGTT

TCTAATCTGCTATTAATTTGTCCATACAGTTCTCCATTTCAGTGACTTAACTTT
TTTTCATTTCTAGAGTTGGATATGCTTCTTTTTCAAAATTTGCTATTCTTTAATT
CATAATATTGGATTTTTTCATTATTATTTTCAATCATTATTTTATTACTCCAAT
AATTTTAACCATATTTATGACCCCTTAATTTTGTTATATTATCTGAAGTTAGTG
GGGTGCTAGTTCTTTTATTTGTTCATTTGCACACTCTCTGTCTTGGTGGTTCCT
TTTCTTATGTAGTTGATCTTTTTTATCTGGGAGTTTATCTTCAGAAGAGGCTGC
ATTTTTTCTAGTGACAGTTCCTTGGGCTGTGGTTAATGAAAGAGTCCCTACAT
AGTTTCAAATTAGATTTTGCTGTGTCCTAGTTGTTTCAATGGCCTTCAAACAA
TTTTACATTATCATCTCAGGTTAGGGCTTTTCTCTTAGGTTCGTAATATAAATT
TGCATCCTAGACCCATGGCACAAAACTTAAGCACAGGGCTTAAATTTTGATG
CCTCAAGTAACTTTTGTTTGTTTTCCATCCAAAGAGTTGGCTAGAAGCAAACT
TTCTTAATATTTAATTAAGGCTTTTGGTATGCTTTTAAAAATCCCCTTTTAAGT
GATCAGGCAGTTCTTTAAGATATCAGGCTTTTGATGATACCTGGAATCAAGTT
CTAGCTTCTTTACATTCTGTAGGCAGAAACCTCATTTTTCCTCCCATGAAAAC
ATAAGAACTCAGCAGATCTACACCTGCACTTATTCCTCAACATCTCCTGGCTT
CATTTCTTTGCTTTGATTTCCTTTTCAATTCTGGGATTGGAGCTTTTCTTTTATT
CTTATAAATTTGACTATGCATTAAAATGTTTATTTTGTGACATTTTACCCAGA
ATTTCAATGGTTTTTGTAGCAGCCAGGAAAGTCTAGCTCTCTGTTATGATTCTT
TTAACTTGTTCAGTCCTGGGTTGTGGAGGTATGAGGAGCTATCCATCACACAT
GGCAATTTCAAAGCAACATGCAAAATACAATAATATAAGCGGAACACATGA
AAGCTGGAGGACAGTGAGAGATTTTTGGTTTTGCATAATAGCTTTACTGAGA
CATAATTAACATGCCATACAACTCACTCATTTATGTGCAACTCAATGGTTTTT
GTATTTACAGAGTTGTACAATTATCACCACAATCTTAGAAAATTTTCATTACC
CCCAGAAGAAACCCCATATCCATTTGTATTCAGTCTCCATTTTGTCCCATTCA
CTCCCTTTAGCCCTAGGCAACCACTAATCTACTTTCTGTCTCTGTAGATTTGTC
TATTCTGAACACTTCATATACATAAAATTACATTCTATGTGGTCCTCTGTGATT
GGCTCCTTTCATGTAGCATAATGTTTTCAAGGTTTATCTATGTAGCATGTATC
AGTATTTCATTCATTTTTATGGCCTAATAAAATCATGCACCACATAATGACCT
TTTGGTCAATGACAGACCTCACACTTGACAGTGGTCCCATAAGATTATAATGC
AGTTGAAAAATTCCTATTGCCTAGTGACATCACAGCCGTTGTAATATCATACA
GTCATAAAGCAACTCATTACCTTTTCTGTGTTTAGGTACACACATATTTACCA
TTGTGTTATAGTTGCCTACAGTATTCAGTATAGTAACATGCTATACACGTTTG
TATCCTAGGGACAGCCGGCTATATAGCATATAGCCTAAGTGTGTAGTAGGCT
ATACCATCTAAGTTTGTGTAAGTACACTCTATGATGTTCACACAATGATAAAA
TTACTCAATGATTAAATTCTTAGAATGTATCCCAATTGTTAAGTGACATATGA
TTGTATTCCATTGTATGGCTATCCTATATTTTATTTATGCATGAATCAGTTGAT
GAACATTTGTGTTGTTTCCACTTATTGGTTATTAAGAAACATGTTGCTCTGAA
CATTTGTGTACAAGTTTCTATGCGAGCATATGTTTTCAGTTCTTTGGGAGATA
TATTTAGCAGTGGAATGGCTGGGTCATATGGTAATTCTATGCTTAACCATTTT
GGGAACTGCCAGACTATTTTCCAAAGCAGCTGCACCATTTAACATTCCTATCA
ACAGTGTATGAGGGTTCCAATTTCTCCATATCCTGGACAACACTTATTATCTG
TATATTTTATTTTGGCCATTCTAATGCATGTGAAGTGGTATCTCATTGTGACTT
TGATTTGCATTTCCCTGATGGCTAATGATATTGACCATCTTGTCATGTTTATTG
GCCATTTGTATATCTTCTTTGGAGACATGTCTAATCAAATCCTTTGCCCATTTT
TAAATTGGCTTATTTGTTTTTGTTAATTATTGAGTTGTAAGAGTTCTCAGAAGT
CTTATGTTTAAGACTTGCAATTAATATATTTAAGATATAATCCCCTTAGATAC

ATAATTTGCAAGCCTTTTTTTCCCCATTCTTTGGGTTGGAGAGGTTTTTTTTTT
AATTGAAGTTCTCTGAGTTACAGAGAAAAGTCACAGAAGTAAAATATAATGG
GACTTTTGAAGTACAGAAATATAAAGCCTCCTATTCATCCATTTAAGTCGCAG
TAGAAACATGACATTTTAGTAAATAAGACATACAAAATACAGTGCAATTTAT
AACAGGACCAGTTGTGGAAGTGGACAGAGAAAAAAAGGAAACAGATGAGAG
AGGTGGAAAGTTAAGAGGAGAGATAAATGCATAGTTCTGCCTTTTTGCTCAA
ACCAAGGATGGCATCATTATATAACACTGGGTGGAAGCCAGGATACCCAGGA
AGAGCAAGACTGAAAATAAATGGTAGGAGAGATTCAAGGAATGCCTTGCTTT
ATATGCCTTTGGAGAGGCAAGGGTCTGGCCCAATGAGGTGGAAAGGGAGTCT
ACAAGGGAAGTGATCAACCAGAGAAGAGTGTGAGAGGGTTCAGAGGAAAGT
ATAGTGGTGGGATATTCGGAGCATTAACCAATGATCCAAAGTCTAGCCTCTT
ATGTTGACAATAATAAAACAGCTTGGGTGGTAGGAGCTAGGTTTCCCAACCC
CAAACTGGAAATGGATTTCCTGTGCTGGGAAAGTTGGAGGGAACAGGTGAGC
ACCAATGTATTATCAAGTCTCTGAGGCCTATATCCCCATTGGTTCTTTGTGTCC
TAACACCATATGCTGAGCTCTGGGATGGTGGCAGCAAACCACCATCGATTTG
TGGAATTTAAAAGCTGACAAAATCTGTTCCCCATCTTTTGGATAGAGCTTACA
GTCACAAGACCCCAGGAAAGACACAGCTGAGTGGTGTTTCTAAGCAAACATA
GGGCTTTGGAAAGTTAGAGAGACTTTACCATGTTTGGTGAGCATTAAAAGAA
GAAATTGCTGCCTGAAGTATCCAGGGGAATAGGTCTCTAAACAGCCTCACAA
TTTTCCACCACCCATGGTAGTGGAAAAACAGTATAATGACTTTCAAGCTCCCA
AGGGCAGCATGGAGGTAGGGAAGGAAACCTGCATGGCAGTAAAGCTCAATC
AGGCTGAGACATGCTTAGGAAATCAAGAATCTGTGGTTGCAGAATAGAACAT
GAATATTATAAACTGAGGCAATTTTAAAAAGGCGGATGGGGAGTGGTGATGA
GAGGAAGAGACACTATTCTTTTATTCTCTTTTTATATCAAATAGTCAATCAAT
ACTGTACCTAATGATTGATTTCAACCTTAATGCTCTTCCTAATTGTTTCCCTTC
AAGTTAGACACCTGTGTTCACTGCTGTATTCTTAGCATCCAGAACAGAACTGG
CTGGCTCAATAAATATTTAATGAATGAACGAATCCATAGGATTCCCACATCTT
AGAGTTTTTGTCCACTTTTTTAAAAAAATTAATGTTGCAGGCTTGATTCCCTG
GGAAGCAGACTCTAATAGGGAGGGTAGTGTGCTTAGCATTTATTTAGGATCA
ATATCTATGGAAGAAAGAAGAAGGAAGCAGGTTTGGGTAGAGGTGGAGGTT
GAGCTACAATGCAGGTTCAAGCAGAGTCTCCCCACAGAGAGCACTGAAGTGA
GGGAGCCCTTCACAGTTGTCCTAATATTGCCAGGCTTTCGTATTCATGCATCG
ATTAGTCATTGACATGGGCTGCCACAGAAAGGGGCAGGTCTTGGATCAGGTG
ACCCTCTGCAATGAAACAGTCTTGAGGAGCCTGACAGCTGGAGGGAGTCTAC
CAACAGCACTCCCAAAAGCTTGGGCAAGACATCTTTCATTGAAGAGGGACCT
TGGCAGCATACCATGGTGTCCACCCCATGAAGCAACTATATCTTAGAAAACT
TAAGTCCTTTGTCTTACACCTTGGATCCAAACTTAGTTTTGTCTGACTCCAGA
GCCCATAATCAGAATAGAGTTTCTCTATTGTACATGGTGATGTAAACTCAGTC
CTAAACTGGCGAGGGTGCTGGTTGGACTTGGGAAGAGTTATTCTCAGTGTTCT
GGGTTAATGCAAAATGCAGGCCAAGGTGAAGGACCTTTGTGTGGCACTGACC
CAATCCATCAAGATGAGTTCATGTCCTTTGTAGGGACACGGATGAAGCTGGA
AACCATCATTCTGAGCAAACTATCGCAAGGACAAAAAACCAAACACCGCATA
TTCTCACTCATAGGTGGGAATTGAACAATGAGAACACATGGACACAGGAAGG
GGAACATCTCACACTGGGGCCTGTTGTGGGGTGGGGGAAGGGGGGAGGGAT
AGCATTTGGAGATATACCTAATGTTAAATGATGAGTTACTGGGTGCAGCACA
CCAATATGGCACATGTATACATATGTAACTAACCTGCACGTTGTGCACCTGTA

CCCTAAAACTTAAAGTATAATAAAACAAAACAAAACAAAAGAAACACCCTA
TGCTCCACTCAGCTGGGAGGCCTGCACAGCAGTGATCTGTATTGAATGATAT
GATTAACTGATAATTGCTACTGTAACTAAGATTACAGTTTGACATTGCTGCCC
ACCTGCCTTTCTACCAGGGTGGTAGTTCACTAAAATATTATACAATCAGATAA
AATGAATTAAACTCAGCTTGAATTGTAGAGTATATTAAAGTGATTCAAATTAT
GGCCATAAGATATTGATTAGATCAGTAGTTTGCAGAATGTGGCCCCCACCCA
ATAGAAACAACATCATCTAGGTACTTTATCAAAATTGAAATTTCCTGGTCCAC
CATATACTTACTGAATCAGAAACGGTGGTGGTGGTAGGGCCCAGGTGATTCT
AATGTATGGTAGAGTTTGAGGAATACTGGACTACATGAAAATTAGTAGGAAA
AAATAAAAATTGGGTTTATAATTAGCAAATAAATGTTAAATAAATGATTCTTT
TAAAAACTATTTGTAAAATGACTACTCAAGTCAAGAAATAGTACATTGCCAG
CACCTTGGATGCCTGTTTATCCCTTCTCACTCACACCCCAAAGCAAACACTTC
GTTTTATTTTATGCTTTTAATATCTAGGAATACATCCCTGAGTACAAGTTTTTG
AAGTTTGTGTTAATAAGCAAAATAGTAATAAATATTAAAAAATGATTGAGGC
CACACGCAGTGGCTCACACCTGCAATCCCAGCACTTTGGGAGGCAGAGACGG
GCATATCACCTGAGGTCAGGACTTCGAGACCAGCCTGGCCAACATGATGAAA
CCCCGTCTCTACTTAAAATACAAAAAAACTAGCCAGGCTTGGTGGCAGGTGC
CTGTACTCCCAGCTACTCGGGAGGCTGAGGCAAAAGAATCGCTTGAACCTGG
GAGGCGGAGGTTGCAGTGAGCCGAGATCAATCGCGCCACTGCACTCCAGCCT
GGCCAACAAGAGCAAAACTCCATCACACACACACACACACACACACACACA
CACACACACACACACACACACACACACGATTGGACTATTTCCTTTACTTTGTTCA
TAGAACTTGTTTTACATAACGGACCTCAGGCTATCCAAATGATCACAACTTCC
TAATTAGGAAGGTCTGAATTAATGAAGGTTCAAGATTGCCTCCTTGGGGGCT
AATGTGTATGCAAGCTGCGACCCACTGGTAACAGCTTTACTATTTACTCTTCC
CTGCCAGGGGATTAGTGGAATCTAAATTGAACAGTTAGGTATTTAAAACCAC
TCATGTGGTTTTAACCACTAAAAGGGTTGTTAAGCAAGTCTTCTTTAATTTTTT
TTTTTGTTGAAAATATTATAAAATTGGTGTCTAAATGCTGACAGTCAATGGGC
AACTTGGGAAATTACTCAATGTCTGTTTTGGGAAAACAGTAACTGGCACTTAC
ATATCACATGTGATCTGAGTAGCGTTAACTCCTTCTCTAGAGTACAGGATTGT
ATGAAGTTGAGGTCCTACCCTTAATTCTTTCACGTTGAAATATATAGTTACAG
GAAAGTGAGTTTAAATTGATTGCATGTTACTGTCTCCCTTTTAGAGTCATTTA
TTTTAAGAAGCACTGAAAGGCCGGGCGCGGTGGCTCACGCCTGTAATCCCAG
CACTTTGGGAGGCCAAGACGGGTGGATCAGGAGGTCAGGAGATCGAGATCA
TCCTGGCTAACACGGTGAAACCTCGTCTCTACTAAAAATACAAAAAATTAGC
CGGGCGTGGTGGCAGGCGCCTGTAGTCCCAGCTACTCGGGAGGCTGAGGCAG
GAGAATGGTGTCAACCCGGGAGGCGGAGCTTGCAGTGAGCTGAGATCGCGCC
ACTGCACTCCAACCTGGGCGACAGAGCAAGACTCCGTCTCAAAAAAAAAAA
AAAAAAAAAGCACTGAAAAAACTTGCGTAAAGGCATCACACTCTGTTGGGCA
ATGGGGAAGGGGAAGGTGGGTTTCTCGCCCTCTGTCACTGATATGGTTTGGTT
GTGTCCCCACCCAAATCTCACCCTGAATTGTGATAATCCCCACGTGTCAAGGG
CACGGCCAGGTGGAGATAACTGAATCATGGGGGCGGTTTCCCCCGTATTGTT
CTCATGGTAGTGAATTAGTCTCACGAAATCTAATGGTTTTATAAAGGGCAGTT
CCCCTGCACAAGCTCTCTTGCCTGTCATCATGTAAGATGTGACTTTGCTCCTC
ATTTGCCTTCTGCCATGATCATGAGGCCTCCCCAGCCATGTGGACCTGTGAGT
CAATTAAACCTCTTTCCTTTATAAATTACCCAGTTTCGGGTATGTCTTTATTAG
CAGCATGAGAACAGACTAATACAGTCATTTACAAGCTATCTGGGGCTATAAG

AAGGATATGCATGTCCTAGGTGAGGAATACTGCACAGGGGAAAGAGTCTGA
GTGAGTAGGGCAGATTGTGCAATGTGAGGCTTTAGGGGATCCCAGGGGACAA
GACTTGGGAGACACATTTGTAGAAAGGTTGGAGTTTTACAGTGATGCCACTG
AATTATCAATGGAAGGATATGATTTCCTTCGTACAAGAAATCTGGAAAGTAT
ATGGCATTTATAGACAGGTTTCCACTTCAAAAAATCAAGCAGCATATTATTTT
GCCAGTTTAAAAGTTAACCCTGCTTGCTTTTTGTTTTGTCTTGTTTTGTTTTGA
GACAGTCTCACTCTGTCACCCATGCAGGAGTACAATGGCGCGATCTCGGCTC
ACTGCAACCTCCAACTCCTGGGTTCAAGTGATTCTCCTGCCTCAGCCTCCCAA
GTAGCTGAGACTACAGGTACCTGCCACCACACCTGGCTAATTTTTGTATTTTT
AGTAGAGGCAGGGTTTCACCATGTTGGGCAGGCCAGTCTCGAACTCCTGACC
TCAAATCACCCACCTTGGCCTCCCAAAGTGCTGGGATTACAGGCATGAGCCA
CCAAGCCCAGCCTGGTTGCTAACTATTTGGATAACTGTTTGGAATCACTACAT
TCCTGAGTGAGTGATTCAGACAACAGAGGTTTTTTAAAAAAACTTTAAAAAA
ATTTTTATCTTAATAGTTTTTTTGGGTACAGGTGGTTTTTGGTTACATAGGTAA
GTTCATTAATGGTGATTTCTGAGATTCTGGTGCACCTGTTACCCACACAGTGT
ATACTGTGCCCAATATGCAGTCTTTTTTCACTCACCCTCCTTCCACCGTTTCCC
CTGGAGTCCGCAAAGTCCATTATATTATTCTTATACCTTTGCATCCTCATAGCT
TAGCTCCCACTTATAAGTGAGACCATATGATATTTGGTTTTCCATTCCTGAGT
TACTTCACTTACTTACAATAATGGCCTCCAGCTCCAACCAAATTGCTGCAAAA
GACATTACTTTGTTCCTGTTTATGGCTAAGTAGTATTCCATGGTATATATACTA
TTTTCTTTCTTTTTCTTTTTTTTTTTTTTTTTGAGACAGAGTCTCGCTCTGTTGCC
CAGGCTGGAGTGCAGCAGCGCGATCTCCGCTCACTGCAAGCTCCACCACCCG
GGTTCATGCCATTCTCCTGACTCAGCCTCCCTAGTAGCTGGGACTACAGGCAC
CCACCACCACGCCCAGCTAATTTTTTTGTATTTTTAGTAGAGACAGGGTTTCA
CCGTGTTAGCCAGGATCGTCTCAATCTCCTGACTTCGTGATCCACCCGCCTCG
GCCTCCCAAAGTGCTGGAATTATAGGCGTGAGCCACCGTGCCCGGCTATATA
CTACATTTTCTTTATCCACTCCTTGGTTGATGGGCACTTAGGTTGGTTCTGTAT
TTTTGCAATTGTTAATTGCAGACAACTGAGGTTTTAATGAAGATTATAGTGTT
GAAGTAGGATATTTCTAATATTCTAGTCTTATGAGGACTTATAAAATTGGGTA
GATTATCAAATCCTCAAATTACGGCATATTCATTTTGGCTTATATTTAAAATA
TTCCACCATCAAGACTGGGGAAAAAAGTTCATCAGAAACATACGCTGATATT
TGGCTATATTGTTTGTTTTTGCATGCATTTATGCAATAAACAAACATCTGATTT
CTTGCACAGTCCCTCAGATATTCTCCTCACATTAAAGATTCCACTTACTTATTC
TGTGATTTCTCTTATTCTATGAGACAAAAATACAACAGAATGTCAGAAGAGC
CAGCTGAAAATATTCCATGTGCAGAAATTTATTTTAAATTTTATTGCATCACA
TTATACAAGCATTAATCATGGCTTCATATTGATGACTATTTAAATGTGAAAAT
TCACTCATGTCAGTACTTTTTGGCTATTTACAAGTAAGGAATTTCTATGTACTT
TATATATCTCTGTATTTGTATGTACATATGCAGGAATACATGACTATACATAT
GTACACACAGAAATACATCTATGGCCATACACATAGCTATAGATATGTCATA
TATAATAATCTTCTCAGAAAGGTCTAAAATTAAAAAAAAGGAAAGAAAAAGT
TGTAAACAGGGTCTTGTCTGTGTTGTTGGTGACATGGAATTAGGACCATATGA
TGACATTCTAGGAATGTGTGTCCATGTGTCTGAATACCCTTTTTAGCCCAGTG
TCTGTAAAATTCACATGGGATAGAATGCAAAAAAGGTAGCAAACAGGCTGA
AAAACAGGCCCAGTATACAAGTTCCCCTTGATTTTAAAAACTTGAGAATGTA
ACTGCCCATAATGTGCATGCTTGCTTGGTCAAGAATGGTCTATGGATAGATAG
CAATTGTGCGCTGGACTGCAACGTGATTTCACAGCAGTGTGACTCCAACGCC

AGTCCTCTACTGGATGCTGCCCCTCCTGCGGCCGCCGGTATCTGCAACATGCA
CTGTGGTTGCATTTTTAGTCATTCACTTGAGTGTGCTTTTGCACAGCCAGAAG
AAGTAAACAGAATTCAGGTCCTTGGCCTGGAAAGGGACTATTTTCTGGAAAG
AGAAAAGAAGGCATAAAGATCTTATGCATACACAATTGCTTTAAAACATGAG
GTGCAATAGTTTGTACTTACCATGCACCAAAGGCATTCAAAGAATTTGTAACC
CCAAATGGAGTCAGGGATGGTATAAACATTTTTATGCTTGCTTTATGGGTGAA
CAGACTGAGTCAGAAGGGCAGTCTAAAGCAACATCACATCCTGCCAACTATC
GGGGAAATAGAAACAGGCAGGCAGATACTATGCTTTCAAGCAAAAATTAGA
AATGGCCTTTTTGGGAAGAAAGCAGCCAGTAGTGGTGGGGCCAGAAGAAATT
CTTCATCAGTTACTGGTCCTGCAGCCCAGGACAGCCCTGTTGCACCTTGGTCT
CTGGCCGCTGAGGCTTTCAGGAAGAATTCTCTTTGCTAATGTCTGCCTCCCAC
CCACCTAGAAAGAAGGGGCATGTTTCCCTGGGGATATTAAAGAAAACAGCCC
GGCTGGAGGAAAAGAGAGAGGACGCTTTGTTTCAGGGAACTCCTAGGCCAAT
TCATGAAGACCTTGACTGATAGGCCAGAGTCTGGCAAGAAGGCAGAGAAGT
GTTGCAGAAGGTATGTGGGAGGGGAAGATCAAGCAAAATTTGCAACGATTA
AAGTAAAAAACAAGCTTCTTTTGGATTGGGAAGCTTCTAAGGAGTTTATTTTG
ATCCCTGTTACTGCAACAATGGTAAGCTTTGCTTCAGGGAGTTTAAAGCACCC
ATTCAACCTCGGAATGGCCTCAGACGCTGAGCACACCTCTCACCTACATGGTC
ACGGCCACTCGTTGTGGTGTGCTTTTCCATTCTTCTTCCTCTCCTTTCGTTCCTT
CTGGAGACGTTCCAGTTCTGCTTCATACATCATCTCCTGAATCAAGTACTTCT
CTCTTCTCATTCGATCCCTTAGGTCTTTTGGGAGGTCTGGGATCAGATATGAA
ATGAGGTGCTTTATACAAAACACGAGGTGCTAAAACAGAGGAGAATGTAAC
AGCGTCAGGTCTACTGTGCATCCTGAATGAGGTTAGAGAGTTGGCTGAAGAA
GGATGTGGACCCTCTGGGGTATGTGGAGAGTGCAAAGTGACCAGGGGAGGC
GTGTGTAATGCAAGATGATAGGACAGGTGAGGCCTGAACTGGTCTGATGGCA
CCACCTAAGGAATAGAAAGAGGAGTCTGGAAGGAAGGTGGGAAAAGAATGG
TCAGATAATTAATTCAAAGGAGAGCCAGAAGAATGTATTCTGTGGAGTCAGA
AAACCCAGGTTGAAGCTGGTAGCTTTCACTGCTTAAGGGCTGTCTGTCTACCT
GAGGATGAATTATTTAATTTATCTGAGCCTCATTCCCCTCATCTATAAAGTGA
GTATTGTGAGAATTATGAGAAAATGTTTAAAAATGCCTTGTAGACTGTCAAG
TGCTTTACAAATCTTGGTTATAAGAATTATCAGTGGGCCTAGGGAAATGATCA
TTTCAGGAAGAAGCTTTTTTCTTTTTGGTAAACAGTCACAAATTCTTTACTTGA
AATACTAATAGCTGGAATTCTAGAGTCACTGTCACTTCTGCTCATCTATATAA
TGGTGACCAAGGAAGTACTAGATGACAATTTAAGGGCTATTACAATTCTAAA
GTATTAAATAGAGACTAAATCATGTTCAAATGATACTGCTAAGTATTCTTCGT
TTTATTTAAAACTTGATAAATTTGGTGATGCATGGATGGCACAAAATTAAGA
AGCTTTTCTGCATAGTTCTCATATTTTCAAATAGTTAATAACAAAATATTAAA
AGAAAATGGAAAATTCTATTTGAAAAGCCAAATAAGGCAGCCTTTTAAAAAG
TTTTGTCTGACAGATTTACAGCAGAATTCTACCAGAGGTACAATGAAGAGCT
GATACCATTTCTATGGAAACTATTCCAAAAAATTAAAAAGGAGGGACTCCTC
CTTAACTCATTTATGAGGCCAGTGTCATCCTGATACCAAAACCTGGCAGAGA
GACAACAAAAAAACTTCAGGCCAATATCCCCGATGAACATTGATGCAAAAAG
CCTCAATACAATACTGGCAAACCAAATCCAGCAGCACATCAAAAAGCTTATC
TACCATGATCAAGTTGGCTTCATCCCTGGGATAAAAAGTTGGTTCAACATATG
CAAATCAGTAAACATCACATAAGCATAAGTAATTCATCACATAAGCAGAACT
AAAGACAAAAACCACATGATTATCTCAATAGCTGCAGAAAAGGTCTTTGATA

ACAATCCAACATCCCTTCATGTTAAAAATTCTCAATAAGCATTCCCCTTAAAA
ACCGGCACAAGACAAGGATGCCCTCTCTTACCACTCCTTTTCAACGCAGTATT
GGAAGTTCTGGCCCAGGCAGTCAGGCAAGAGAAATAAATAAAGGGTATTCA
AATAGGAAGAGAGGAAGTCAAATTATCTTTTTTTTGCAGATGACCTGATCCC
GTGTCTAGAAAATCCCATCATCTTGGCCCAAAAGCTTCTTAAGCTGATAAGCA
ACTTCAGCAGAGTCTCAGGATACAAAATCAATGTGCAAAAATCATTAGTATT
CCTAACCACTAACAACAGGCAAGCAGAAAGCCAAATCATGGATGAACTCCCA
TTCACAACTGCTGCAAAAGAATAAAATACCTAGGAATACAGCTAACAAGAAA
AGTGAATGACTTCTTCAAGAACTACAGACCACTGCTCAAGGAAATCAGAAAG
GACACAAGCAGATGGAAAAATGTTCCATGCTCATGGATAGAAAGAATCAATA
TTGTGAAAACGGCCATCTGCCCAAAGTAATTTATAGATTCAATGCTATTCCCA
TTAAACTACCATTGACACTCTTCACAGAATTAGAAGAAACTCTTTTAAAATTC
ATGTGGAACCAAAAAAGAGTCCAAATAGCCAAGACAAGACTAAGCAAAAAG
AACAAAGCTGGAAGCATCACACTACCCAACTTCAAATTATACTAAAAGGCTA
CAGTAACCAAAACAGCATGGTACTAGTACAAAAACAGACACATAGACCAAT
GGAACAGATTAGAGATCTCAGATATAAGACCACACATCTACAACCATCTGAT
CTTTGAAAAACCTGACAAAAACAAGCAATGGGGGAAAGGATTCCCTACTTAA
TAAATGGTTTTGGGAGAACTGGCTAGCCATATGCAGAAAATCGAAACTGAAC
CCCTTCCTTACACCTTATATAAAAATTAACTCAAGATGGATTAAAGACTTAAA
TGTAAAGCCCCAAACTATAAAAATCCTAGAAGAAAATCTAGGCAATGCCATT
CAGGATGTAGGCATGGGCAAAAATTTCATGATGAAAACACCAAAAGCAATTG
CAACAAAAGAAAAAATTGACAAATGGGATCTAATTAAATGAAAGTACTTCTG
CACAGCACAAAAAAACTATCATCAGAGCAAACAGTAACCTATAGAATGGGA
GAACATTTTTGTAATCTATCCATCTAACAAAGGTCTGATATCCAGAGTCTACA
AGGAACTTAAACACATCTACAAAAAAAAATACCAAGCAACCCCATTAAAAAGT
GGGCAAAAGACATAAACAGACATTCTTCAAAAGAAGACATTCATGCAGCCA
ACAAACATATGAAAAAAAGCTCAACATTATTGATAATTAAAGAACTGCAAAT
AAAAATCACAATGAGATACCATCTTACACCAGTCAGAATGTTGATTATTTAA
AAGTCCAGAAACAACAGATGCTGGCAAGGTTTCAGAGAAAAAGGAACACTT
TTACACTGTTGGTGGGAGTATAAATTAGTTCAACCATTGTGGAAGACAGTGT
GGCAATTCCTCAGTGATTTACAAGCAGAAATACCATTTTACCCAGCAATTCTA
TAACTTGTATGGCAAAGGACACAAACAGATACCTCTCAAAAGAAGATATACA
AGCAGCCAAAAATAATATGAAAAGATGCTCAGAATCTCTAAGGAGATTAGA
GAAATGCAAATCAAAACCACAATGAAATATTATCTCATACCAGTCAGTATGG
CGCTTATTAGAAAGGAATATAAATCATTCTATTATAAACATACATGCATGCAT
ATGTTCATTGCAGCACTATTCACAATAGCAAAGGCATAGAATCAACCCAAAT
GCCCATCAATGATAGGCTGGATAAAAAAAATGTGGTACATATACACCATGAAA
TACTATGCAACCATAAAAAGGGATGAGATAATGTCCTTTGCAGGGACATGGA
CAGAACTGGAAGCTGTTATCCTCAGCAAACTGACAAAGGAACAGAAAACCA
AATACCACATGTTCTCACTTATAAGTGGGAGCTGAATGATGAGAACACATGG
ACACATGGTGGGAAACAACACACAAAGGGGCCTGTTGGGGGTGGGGGTGGG
GGAAGGGAGAGCATCAGAAAGAATAGCTAAGGGATGCTGAGCTTAATACCT
GGGTGACGGGTTGATCTGTGCAGCAGATGACCGTGGCACACATTTACCTGTG
TAACAAACCTGTATGTCCTGCACATGTACCCTGGAACTTAAAATAAAAAAAT
CCCAAACAAACAACAAGAAGTTTTGTCTGAAAAATTTTAAGACAGTGATGGG
TTAAAAATATCTTCTTTAAAGAGAGAGTGTGCCATTGGTAAGATAATTTCCAG

GGAAGAGCAGCTTAATATTTTCTTCTTTTGGTTCCCTGGACTGAAGGAAAAAG
CAGTGATAGAATAGTCTTGAAGAGGTCTAGGAAACTTTAGATCCAAGCCGAG
AGGCAACCTTGGCTTTTATTAATCTGGCTTTAATATATGTGACAAGAGATGAA
ATTTCCACTTATGACTAGAGTCATAGAAATGCAAACTATTTTTACAGCAATTT
TCTTAAACCCTGAAAGAAAATAGATAATATATTTTTATTGACATATAATATAG
ACAATTGCTAGACAGATTTTACTTTATAATTCCATTTTGAGTCTTAGCTAATA
AATACTTACTTGGATGCTTGAATATAAATAATTCCGTGATGATACACAACCAG
AAATACCACTTGTAATACCCAGTCCTATTGGAAAATGCTTATGATCATGTTGG
GAAATCCTGATAACTATTTGAAAATCATAAATTAACTATATGAATCAGTATAT
TTCAATATGTGTGGTAATGATGATGGCAATAATTGGAGACATTTAGAAAGAG
TGAATCTCCAACTTGACAAAAACAAGCAGTGGGGGAAGGATTCCTTGTTCAA
TAAATGGTGCTGAGATAACTGGCTATCCATATGCAGAAGAATGAAACTGGAC
TTCTACCTATCATCATAAACAAAATTTAACTCAAGATGAATTAAAGACTTACA
TGTAAGACCTCAACTATAAAAATGCTAGAAGAAAACCTAGGAAATACCCTTC
TCAATACTGGCCTGGGCAAAGAATTTATGGTGAAGTCCTGAAAAGCAATTGC
AACAAAAACAAAAATTGCTAAGTCAAATCTAATTAAAGAGCTTCTGCACAGC
AAGAGAAACTGTCAAAGAAGTAAACAGACACCGTACAGAATGGGAGAAAAT
ATTTGCAAACTATGCACCCAATAAAGGTCTAATACCCAGAATCTGTAAGGAA
CTTAAACAAATCAACATGTAAAAAACAAATAACCCCATTAAAAAGTGGTCAA
AGGACACAAACAGATACTTCTCAAAAGAAGATATAGAAGCAGACAACAATA
ATATGAAAAAATGCTCAGAATCTCTAAGGAAATTAGAGAAATGCAAATCAAA
ACCACAATGAGACACCATCTCACACCAGTCAGTACGGCTTTTATTAGAAAGT
CAGGCCAAGTGCAGTGGCTCACGCCTGTAATCCCAGCACTTTGGGAGGCCAA
GGCGGATGGATCATGAGGTCAGGTTAAGACCAGCCTGGCCAAGACAGTGAA
ACCCCGTCTCCACTAAAAATACAAAAATTAGCCAGGTGTGGTAGCGGGTGCC
TGTAATCCCACCTACTCAGGAGGCTGAGGCAGAGAATTACTTGAACCTGGGA
GGCAGAGGTTGCAGTGAGCCGAGATTGTGCCATTGCACTCCAGCCTGGGCGA
CAGAGAGAGAATCAGTCTTAAAAAAAAAAAAAAAAAGTAAACAAATTAACAG
ATGCTGGCGAGGCTTCAGAGAAAAGGGGACACCTGCACACTGTTGGTGGAAA
GGTAAATTAGTCCAACTACTGTGTAGTCTGGAGATTTCTCAAAGAACAAAGG
GCTTAACTACCATTCAACCCAGCAATCCCATTTCTGGGTATATACTCAAAAGA
AAATAAAGCATTCTACCAAAAAGACACATGTACTCATATGTCAATCACAGGA
CTATTCACAATAGCAAAGACATGGAATCAACCTGGGTGCCCATCAATGGTGG
ACCAGATAGAGAAAATGTGGTATGTATACACCATGGAATACTATGTGGCCAT
AAAAAAGAATGAAACCATGTCCTTTGCAGCAACATGGATGTAGCTGGAGGCC
ATTATCTTAAGTGAAATAATGTAGAAACTGAAAACTAAACACTGCAAGTTCT
TATTGTACGTACTTAGAGTGGAAGCTAAACACTGGGTACACATGGACATAAA
GATGAGAACAACAGACACTGGGGACTACTATAGAGGGGAGAGGGGGAGGGG
ACAAGGGCTGAATAACTACCTATTGAATACTATGCTCGCTACCTGGGTGGTG
GGTTCAGTCGCACCCCAAACCTCAACATCATGCAATGTAACTTTGTAACAAA
CTTGCACATGTACCCACTGTATCTAAAATAAAAGTTGAAAAAGAAAAAGAAA
AAAAGAATGAATCTTACTGGGCTGATGTTTTCCAAATGTTTCCAAATTATGGT
CCCAGAATGCCTATAGTCAGAATCATTACAACGAGCTCATGAAAAACAAGCC
ACAGACATGGTATGTCAGAATATCTGGGAGTGAGGATGGAAAAACTGCATTT
TTGCTAAAGTTTAAGAACTATGGCCCCAGAAAATTACATTTAGGAGCAACGA
GTTCAGATTTGTCCATTAAAAACCAACTATGGACTTTCAAATATGTCTTGGAA

AAATCAAATGCTTCAATTAAGAAAAAAGATCAAAAAGAGGTTACATAATATG
ACACTACTACTTCTGGAAATATAGATAATCACATTCAAAATCCCCATCACTTA
TTTTTATTCATTACTCCTACTGACTTCATTATGTCAAAGGAAATAAAGAATTT
GGCTGAGGAAACTTACCTCAAAGACAATGATAAAAGCTAATCGAGCAGCTAG
GACATGCCAAAACTGCAGTGTGTAGCCATAGGGCACCAGTGAATGAGGCGG
GTCACGGTAGTCCCGGTATCTATAAAGACACAGAAAAATGTTGCATTCAAAC
TATGATTTTTCTGATTCTCAAATCTCTTCCATTCTCATTATATAGTCCTCATTC
AATAGGAGGGAAATTTGAAGTTTCTCTATACTAGATATCTGTTGTTTTGCTTG
TCCAGCATTTTTCTTCTGATAAAATAATATCTCTTCTCTGGGGAATCATCTGTG
TGGGTTGACTAGGACTTCCCCCTCCACCTGCCTGATCTCAGCTCAAAAGATGA
GCCAGACCTGATCAGTGACATCATCAAGATGTCTGCCAGAACTATCACTTAG
TGAATGATCACTAAGCCCACAGAAGGCAAAAGCCTGGGACTCCTGGTGGCCT
TGTATGGAGAGAGAGAGCCTGCTTGAGAAGGAAGTCAAGAAAACAAAGCAG
AGACAGCAGAGAGGCAGCATCCTGCTGACACTGCTTGAGGCCTGGGATCCGG
CCATTACTGAAGCTCATATCACCCCTTGACTCTCCCAGCTAACTGAACAATTT
TTTGTTGTTTAAGTCAGTTTACCCTGGATTTCTGCCACTTGCCACTAGAAGAG
TCCTTCCTAATGTATCCTCCAGCTATCACTTTTGAGTCGGCATAAAATTCCAC
AAATGTTTGCCGAATACCTGTTGGGGACCAGGCACTATTCAGAGAACTTTAA
TCCATGTGACTCCATTAAAGTAATACTTTAATAATACAACAAGCTCTCATAAA
TCCACTATGAAGCCCAAGAACTAAAACATTACCAGGAACTTACATCAATTGC
TTCCCTTACCCTGAATTTCATTTTTATTTTTAACTTATATGTATATGCATTTGTC
TGAATATTGTACTATTTAGTTTGTTTTTGAACTTTATAGAAAGGGTACTATACT
ATATGTAGTTTGCTGGGGCTTGCTTTTTTTACTATTATGTCACTAAGATTCATT
CACATTATTGCATGTGGTTTATTTTTATTGCTGTATAATAGACTCTAGTGTGAC
TATATTATGGTTGTTTTATATAGCATATAAGTATACACTTACAGGCTGGTCAT
TCCTATATTTTCTTTTCTTTTATGTTGAGACAGAATCTCGCTCTGTCACCCAGG
CTGGAGTGCAGTGGTGCGATCTTGGCTCACTGCAATCTCTGCCTCCCAGACTC
AAGTGATACTCCTACCTCCCAAGTAGCTGGGACTACAGGTGTGCACAAAAAT
GCCTGGCTAATTTTTTTGTATTTTTGGTAGAGACAGGGTTTTACCATGTTGCCC
AGGCTGGTCTCAAACTCTTGACCTTAAGCTATCTACTCACCTCAGCCTCCCAA
AGTGCTGGGATTACAGGGTGAGCCACTGCACCTGGCCTCTATATTTTCTTTCA
CTTTCCCTAACCATGGAAAGCTTTGAAAAAGAAAGCTCCATCAATGATAGAC
TGGATTAAGAAAATGTGGCACAGATACACCATGGAATACTATGCAGCCATAA
AAAAGGATGAGTTCATGTCCTTTGTAGGGACATGGATGAAGCTGGAAACCAT
CATTCTCAGCAAACTATCTCAAGGACAAAAAACCAAACACCGCATGTTCTCA
TTCATAGGTGGGAACTGAACAATGAGAACACTTGGACACAGGAAGGGGAAC
ATCACACACCGGGGCCTGTCGTGGGGTGGGGGGAGGGGGGAGGGATAGCAT
TAGGAGATATACCTAATGTAAATGATGAGTTAATGGGTGCAGCACACCAACA
TGGCACATGTATACATATGTAACAAACCTGCATGCTGTGCACATGTACTCTAG
AACTTAAAGTATAAAAAGATAAATAAATAAATAAAAAGAAAGCTCCAGTCC
ACTCACTTCAGCAAATTCCCTTTGGGTAAAAGTTGGCTGCAATGCTCTGTGTA
CCACTTCTTTCTGCCTTAGAATACCAAATATTCTTCCCAGGTCATTAATGTATT
TGAGAAGGTGTTTTTAAATTTTTCTCTACATGTTTAGATATTTTCAGCCCGCAA
GCTGATCTGGGTGCCTCACCTGCTATATGACTGAAAACAAACCTGAAAATAT
TATTTTACATGATTTTAATTAGAAATTTTAAAATGTTTTCTTTGGGGGACTTTT
AAGAAAAATGCTAAACAAAACAATCATAATTTAACTTAATGGTGCCAGCATT

AATACCTCCATTCACATTAATAATTTCTGGAATATACTTTTATGATACTATGT
GAGACAGGCAGGAAGAATTATGAGCCAGGAAGCAGAAGAGGAGAAAGACTC
TGTATTTTTATCTTATTTTCAAATGATTACAAAGTTTTTAACTCCAAATAATAG
TTGGCATTTTATTTTGAAGCATTTCATACTTAGGATCCTGTATATGTCTAACAA
TGCTTGCTAAATAATTTCATGTACGAAATGGTCTGGACAAGCAAGTGGGAAA
TCTACTGAAGAAAGTATCAATTGTTGCTTAATTGCTTGCTTCTCCACTAGACC
CCAACAGAGAGCGGCACTAGATTATATTCCAGCCAGTCTTTCTCCCCATCTTA
CTTGTTACTCCATGTGGGAAAGGAAATAAGCCATCTGATGAGATGTGACAGC
CAATGAGACCTGGGAACTTGGCATGAGTGTTTAAAGGGGTGGAGGGTATGCT
GTCACTGCACAGAAATGGGGAGCCCCTGGCTTCTCTCACCTTCACCTTCTGTA
AGATGAGGGACTGAAGTGGATCATTACAAAAGTCCCTATGTCAAAGCATCTA
GGACTCCATGGCTGTTCTTCCTTGTGGCTTTGATTACAGGGAGAACATAGAAT
GTCATGATAGGGAATCTCCAAAATTCTTTGTTTTGTTTTTGAGACAAGGTCTT
GCTCCATAGTTTGAATGCAGTGGTGCTATCATAGCTCACTGCAGCCTTGACTG
CCTGGACTTAAGTGATCCTCCCAACTTGGCCTCCTCAAGTGCTGGGATTACAG
GTGTGAGCCCACAAAATTCTTATAACGTCAAAGTATTAAGAGAAAAAGGTGC
CATCTGGACTCTTCAAGAACTGTGCACAAGTAATCTCATTGAGCTTTTAGTGA
ATGGTTTTATCATTTGTCAGATGAGATAATATTGCCTAACTTAGAAGGCTGAT
GTGACGATTAAATGGAGTAACAGTTTCTGGCACAGAGTAGGTGCTCAACAAA
TTCTCCTTCATGTACGTTTGTATAGATACACATCAAGAGGCACAGTTTAATTC
ATTTCCTTTTAGAGGGAGAGAAAGAAAGAGAGAATGAGCACATGTTGCAGA
GGCTACTGTGGTGGAACCTGAAATGAACTGGGCCTCTTTCCACAGGCTTCTGA
GTTGTTTAACATTTAGAAATGTGATGTGATATACTCTGGGCAAGAGAGAGAA
GCAGAACAAAGACAAGTGCCCTCAACAGCTGTCGCTCAACCCGCCTGGGGAT
TGTCTGTCCCTGAGTCCCTGGAGCAGCGGTATAACCTTCACTACTATATTTTT
AGTTTTTCCCTTTAAGCTAGCAGTTCTCAATTCTGACTGCACCTGGTTTCTGAG
TTCCACTCCCTGACATTCTGTTTTAATTGGTTTGGGTTGTATCCCGGGCAAGA
GAAGGCTTGAAAATTCTTCATGTGCTTTTAATGCATTGTTGAGATTGACAGTC
ACTGCTTTAAACACACTCAGCCAGAAGGGGCCTTCAGAGTAAAACCTCCCCA
CGGTGGAGCCAGCCCAGCTCAGGCAGAGCCACGCTGGGTGCCACACTCACCT
GCAGTACTTAAGAGGAGTCCCCGAGAACTCACTGCCATCAGATTCAGGCTCA
GATCGGTTCTCAAAGTCAGAAATTCGAAATACAGACAAGCTGGCATTCACAT
AGCCAACCATGCACCTATAAGAGGAGGCACATTCTACCATTAGAACACTCAT
CACCTTCTCTGTATCCAGTTCCCTGAAGACCCTGCTTCCCAGACGTCCCTGCA
ATTGTAAATGCCTTTAACTGGCCCTTCCTTGTCAACCAGGACTCAGGTTGGGA
TACGCTGGAAAGCCCTGGGCATGTTGAATCAGGAACAAGTTGCTTGTTTTGGT
TCTCATCCTTTTTCTGCAGAATCATCATATAGAACATCTCTACCCCAATTAAT
GTTTCAGTGAAATAGGTAAAAAGCAGCATTACATGATGGGCATAAATTGTAT
GTAAGCCACAGCTTCTTTTGTTGATGATACACATAAACACAGTTCCTCATTAA
TGGTTTCATATATATTTATAGATATTCTCAAAATGTTCCAAATATGATTGGTG
TGCATATGATCCCCCCAACTCATATCCCATGCTACAGTCCTGTTAGACTATAT
TCCCTGAAAGCTTCTGATGTTTTCATGCCTCTGCTTGTTATCTTTTCTGTTAGA
ATTCCCTACCTCTTCCTTTACAAGTGGAAATAGTATTTATCCTTCAGCGGCCA
ACTGAAACACCACTCCTACTAGGAAGACAACCTAATCACTCCAGTCACAACA
TCTCACTCCTTTCCCTCTCTCTGTAGTGCCACCATACTTACTTAGTTCACCATT
ACCAGATCGCTTGTACTAGAGTTGGCAGCATATATGTATGCCCTCTCCTTTGG

ACTGCAATTTTTTGGGAGCAGAGCAGGTATCTTAGGTATTTGTGTATCCTATG
ACTTAGGACAGTCTTGTGCAGTGTAGGAATCCCATAAATATTGAATCGATCTG
ATTGGAGGAAACTAAAGTATGAGATTCTATTGATCATTCTAGATTGAATGGG
TAGATAAAATCAAGATTTAATCATTCTGTCAATAATTCCATTCAAGAAAAAG
CCCATTTTAAAAGTATCATAGCAAGTAGCACTCTTAGAAACATCTTTTCTCCA
AGTCTCTAAACAGTTTTACATATTGAACATTGTGGTTAACACATGCACTTACT
GAAATAACAATGTTTGAAACTTTGGGTGAAATTAGCACTTAGATACATCCTGT
ATACATTTTTCTTTAAATAGTGTTTTACATACAAAAGCTTGGGTTTTGGGGTC
AAATCTTAGTTCTGCTGGTTTCTAGTCATGTAATATTGGCACGTTATATAACTT
CTTACTTCTTCTTTTTTTTTCTGAGACAGGGCCTTGCTCTGTTGCTCATTGCTG
GAGTGCAGTGGTATGATCATGGCTCACTGCAGCTTTGACCTCCCAGGCTCAA
ACGATCCTCCCACCTCAGACTCCTGAATAGCTGGGACTACAGGCATGTGTCA
CCACACCCGATTTACGTAACTTCTTAAAGTGACAGAAGTTCCCACCTCATAGG
GTCCTTGAGGGGATTATATTAAACAATGAATGTATAGCATGTGATAAGATCTT
AGCCTATAAGTAAGATAAACATGAGTGGTGATTATTCATTTTATTTATTTCTT
GCATCTAGCAGGGGGTCAATAAGTAATTACTGATTGAATGAATAAATCATGA
TCTTCAGATTCTCTTCATTAACAAGTCTGGCATGATTATTCCAAGCTGCTGAA
CACTAAACCAAAATGATGACATGAACACAGTTCTGAGAACACATGGGTTCTG
CATGTTATCGTGCTCAAGACAAATCACTGGCTTTATCATGGCCTTAAAAAAAA
TGAACCAAGCTGAAGGACCCATTTGTCAAGTCACTCGTGGGTGAAAACAAAT
GGTCAGGGAGTGTGTAACATGCCTTAATCCAACTGGCAGCAGGACTCCCCTA
CTGAAAGGGTAATTTGGAATGATGTCTCCCTTACCTAGGAACTGTCTAGACCT
GGACATGAAAATAAATTAGTGCCTTGAGGATGATTAGCTGTCACAGAGAGGG
AGTAAGTTTTAGCAGAGTCTTTCTAGAGATCTGAGCTGTGGGAAGGGAAAGG
TGTTCTGGGAACTGGTGAAAAAAAAAGGCCCAGCACTGAAATGGAATATTCA
TAACCCCTCCTGCACTCTTATCTCACAGACCAGCCCCGAGGCTTTATCAAGAC
TAACGCTACTGTCCATTGGTTGTTCTGCTCAAAAACTCCATTAAGAAAATTTA
AATTTAAATAAAACACCTGTTTCATGGACCTAATGCATGAATGACAGTCATA
AATCCAAGTTGAAAATGTGTCTTTAGAGGGTGGTTACTCCTGAGTGCAAAGC
ATTAAATGTGTACCTAGTACAGGAACACACAGTTATCAGCTGTGTACATCTTC
CTTCCATATCATTTCCATAATGGGGGAAAAGATGCACGGACAGCTAGACGTT
TCACTAATTCCTTTCACCTTTTGTAGACTGTTAGCTATTGTCAATATCAAATAG
TGAAAGTTGTCATCAATTCAGGCAAGTGAGATTTATCACATTTTGCTGTATAA
AATTCAGCTAGTGGCATGCAAGTGTCAACTCACTGACCCATTGTTATAGCAG
GAAGGTCACAGGTAGTAGAGAGCCCATGTCTATCACTGGGTTTACTGGCAGG
AATGCTATATACCCTCACTGTGACCCACCACCAGCTCTCAATGAGACAGAGG
GAGATGAGGCTACACTGATGATCAGGGAGAATTGAGAAGTACCAAGAAAAA
AGTAACAGACGAAGGGTTTCCTGTGAATATAAAGTCAGTTTTCTTAAAGGAA
TGGTAGATGGCCTCTGTAGGCACTTCCAAGAAAGATCTAGAAAATGTCTTCT
ACAGGGTAGTGCCATTTGGATTAGGATGCATATCTCTTCCATGTAACTTCTGA
GAAGACCACACACTCATTTAAATGTGAGAATTCTACCGTGAAAGCCACTCAC
CACTTATAATTTTCAATCCTTTAATACATAAGAAAGGGATGAGCCAGCAATC
ACTGTACTACCCTAAAGGCTTATGACTTTCTATACCTTGACAATTCTGCTACT
CTTATGAGTTCAATAATTTCTAAAGCATTTCTAAATTGGAAAAATAAAAAAA
ATCTGAAAAGTAATAGAGTGGCCAGTGAGGCAGAAAATAAGATTCATGAAG
AAACATACATAAATTTAACTACTTGAAAACATGAAAAATGACTAATGGATTA

GAAAACAAAAGGCAGATCCATAAATTGAAAGTGGCAGAAGGCTCCTAAGTC
AAAGAGAAAGGTTAAAAAAGAAAGAAAGAAAAGAAAAGCTCCAAAGAGCC
CAGTAAAATTGATACAAGGAGGCTATGAGATACACTTTAATCACTGTGGACA
TTGTCTTAAATCTACCAGGTTGTTAAAATTCTAAATCTTTGATTCACTATCAAT
TGTGCATAAAGGTTCCCATTTTGTTTTTGGTTCCCCCTAAAACAAGATAGATT
TCACCATATTTGAATTTGGGAAAGAACTGAAACACCATTACGACGGAAAATC
CAGGGGGTTTACTGTTAAGGGACAGCCTCAAGGGCTGTGAGGTGCGGGATGG
AGAGTGAGTCGAGGCTCCACCTTCACATGGGCTCCCAGAGTTACTATGCCTTG
TGATTCTGGACTTAACCTCTTTTAATCTCTTCCTGCTCCTCAAAGTGAAGATTA
TAATGCACATCCCACTCCCAGGGCTGTTTTGAGAATCAACTAAAAACACATG
TGAAAGTGGTTTGAGTGAAAGTAAGCTTTTTCAGGCTAAGCCAAGCAGAGAA
TGTCAACGATAGTTGCAATTAGAAAGAAATGCCTGGGCAAAATAACATGTCT
TTACATTTCTCTTAAAATGGGATGGTATGCTCACTAAAAGTTGCTAGATAATT
AGTGTGTCATGGCCACATGCTTATAGTAGTCAAGAACTTTACAGAAGCCTTTT
CAAATAACTGGATTGTGAGCTTTATGACATTCCTCAATCATCCACACCTTAGC
CCATTTATGTAGGATTCATAATGAATTGGACCCAGGCCAGGTTTGCATGAGTT
CAGTTGATTTGCTTTGGTGGCACTTAACTCCTTGGGGTCGATCAGCTACAATG
GCTGTAGAGGATGAATCAATCTGGCTCAGGCTAGTAAACCAGGGCTGCTTGT
ATAAGTAGACAGCAGCCCATGTAGGGTGATTTCTGGGTTATTAAGAAGACCC
TGTGATCCCCAAACACACCTCACTCTGACCTATAGACATCAAAATCCCATAAT
AAGAGTTTAACTATTTTGGCTTTAACCAAACAGGGATTTTTAAAGCATGTTAA
GACTTCTTGCCCTTTCTCCAAGGTACAGGAATGAAACTAGGGGATGTTATAAT
TTTGACTTATCTTTATCTGACCATATATAGCGTGCCCTGGCTTTGCAAACCATT
GAGGCAATGCCATAGAATTGATCCTCCCAAGGAGTTTAGTCCACTCAGCTTG
CTTGGGGTAATATAGGCATTGGTCAGGTCAATACCCATGAAAAAGTCATTCA
CTTACTTATTCCACGAATACTTATTGAGGACCTACTGGGTACTCAGTACCTAG
TTCCGTATTTCCTATGTAAGATCTGCAGAGTTTACACATAATGACACAATGTA
AATCAGATTGTAACTGTCCTTTTGCAAACTCATAAAATGATATAACTATATAT
ACATCATATGCTACTGTTTTGGTATACACATTAGCATATCACACATTTCTAAA
TTCATAGTGGGCAGAGGTCAGGGGTGGGGAAAATGATTAATTGCCACTCTTA
CATTGGATTCCATCTAAAACCTGTCTGACTTGTATGTCCTCCAACTCTTTGGTA
GATGAGGACATGTTTTCTTTGGCTGACAGTGTTATTATTATTATTTATAATAA
ATCATAATTCCACCTCCCCTATCACAGGCCTCTCCTTCTGGGCATGTATTCCA
TATTCTCCAGTGCAGTCTTCTGTCTGAGTGTCCAACCTCAAAAAATGGCTAGG
AAATAGAAACTGTATAGTGGTTTTATAGCAAACTTACTTTTGCCCAGCTTCTC
CTTGGCCTGCACAAGGTCCATACTTATAAGCATACACCAAGCGAGGGATAAA
GTCAGATGTTATCGCTATGACAAATGCATTTGTGATAACAGAGAGAATTCCA
ATGCCTTCAAGAATTC

Huil mRNA Sequence from normal colon tissue sample 612N. The start codon in indictaed in green and the stop codon in red.

GTGTAGATTCGCTGGAAGCAGCTGGAGGCTCCAGTTCTCATCTGCTCAGGTGT
CCCCGGCGCCTTGGCGAACTCGGCCACTCCAGTTCCTCACGTGGTGAGCACTC
AGGGCAGCGGGTCGATTTTCCGAGGTCCCATACCTGGGTTTGAGGGGCGCGG
CTCGCAGCGGCGGGTGCAGGGGCGACTGCCAGCCCTCACCCCGCCTCGGGGT
GCGTTCGGAGGCCGACACCTGGAGGACGCCTCCAGTCCCCGCGGGACGCCAC
GCCTGCGCGCCAGGGATCCGGGATAAGAAGTGCGCGCCGGGCTCCGGCTGCG
CGCCGCGGGGCCACCAGTTTGCGCGCAGGGCTCAGGCGACCGTGCGGCCATG
GACACGCCACGGGGCATCGGCACCTTCGTGGTGTGGGACTACGTGGTGTTCG
CGGGCATGCTGGTCATCTCGGCCGCCATCGGCATCTACTACGCCTTCGCTGGG
GGCGGCCAGCAGACCTCCAAGGACTTCCTGATGGGCGGCCGCAGAATGACCG
CAGTGCCCGTGGCGCTGTCCCTCACCGCTAGCTTCATGTCAGCCGTCACTGTC
CTGGGCACCCCCTCCGAGGTCTACCGTTTTGGGGCCATTTTTAGCATCTTTGC
CTTCACCTACTTCTTTGTGGTGGTCATCAGCGCGGAGGTCTTCCTCCCGGTGTT
CTACAAACTGGGAATTACCAGCACCTACGAGTATTTAGAACTTCGATTTAAC
AAATGTGTTCGTCTCTGTGGAACAGTCCTCTTCATTGTTCAAACAATTCTGTA
TACTGGAATTGTTATTTATGCCCCTGCCCTGGCTTTGAATCAAGTCACAGGAT
TTGATCTGTGGGGCGCGGTAGTGGCAACGGGGGTGGTCTGCACATTCTACTG
CACACTGGGTGGTCTTAAAGCAGTTATCTGGACAGATGTTTTTCAAATTGGGA
TCATGGTGGCTGGATTTGCATCCGTGATTATACAGGCTGTGGTGATGCAAGGT
GGAATCAGCACTATTTTAAATGATGCCTATGATGGTGGAAGATTAAATTTCTG
GAATTTTAATCCTAACCCTTTGCAAAGACACACCTTCTGGACAATTATTATAG
GAGGGACCTTCACATGGACCAGCATCTACGGTGTCAACCAATCCCAGGTGCA
GAGATATATTTCTTGTAAAAGCAGATTCCAGGCAAAACTGTCTCTCTACATCA
ATCTTGTGGGACTCTGGGCAATCCTCACATGCTCAGTGTTTTGTGGGCTCGCC
CTATATTCCAGGTACCATGACTGTGATCCTTGGACAGCCAAGAAAGTGTCTGC
ACCAGACCAGCTCATGCCTTATTTGGTACTGGACATTCTGCAAGATTATCCAG
GACTTCCTGGACTTTTTGTGGCCTGTGCTTACAGTGGGACATTAAGCACAGTG
TCCTCCAGTATTAATGCCTTAGCAGCAGTAACTGTGGAAGATCTAATCAAACC
TTACTTCAGATCGCTCTCAGAAAGGTCTCTGTCTTGGATTTCCCAAGGAATGA
GTGTGGTGTATGGAGCCCTGTGTATTGGAATGGCTGCGCTGGCGTCACTTATG
GGAGCTTTGTTGCAGGCAGCACTCAGCGTATTTGGTATGGTTGGTGGACCACT
TATGGGCCTGTTCGCTTTGGGCATTTTGGTTCCCTTTGCCAACTCAATTGGAG
CACTTGTTGGTCTGATGGCTGGATTTGCCATTTCTCTATGGGTTGGAATTGGA
GCTCAAATATATCCTCCACTTCCTGAGAGAACATTGCCATTGCACCTTGATAT
CCAAGGCTGTAACAGCACCTACAATGAGACAAATTTGATTACAACCACAGAA
ATGCCATTTACTACTAGTGTTTTTCAAATATACAATGTTCAAAGGACTCCACT
GATGGATAACTGGTATTCTTTATCATATCTGTACTTCAGCACTGTTGGAACTT
TGGTAACATTATTAGTGGGGATACTTGTCAGTTTATCAACAGGAGGAAGAAA
ACAGAACTTAGACCCCAGATATATACTAACCAAAGAGGACTTTTTATCCAAT
TTTGATATTTTTAAGAAAAAGAAGCATGTTTTGAGCTATAAATCACATCCAGT
GGAAGATGGTGGANCTGATAATCCTGCTTTCAACCACATTGAATTGAACTCA
GATCAGAGTGGCAAGAGCAATGGGACTCGTTTGTAAGCTGCTCTGATACTA
GATATCCTTAAATGATGTTTCAATTTTATATGTTTTCTAAGATAATTGGATCA

Figure 2

page 1

GGTTTTCTTTGTGTGTGTGTGTGTGTTGTATCATGAGTGTTTGGGGGATAA
GTTTTTGTTAAAACAAAGTCTGGACTATCTTCATTTACTACATCATTAATTGAT
GTTACTCTGGAGTTTAGAATTCTGGCATTGACATTTCCCTCTCTTTCCTTTATT
TCGATGAAGCTATAATTGTGAAAATTGTAACTACATAGATGCTGAAAGGCTA
ATACACACATATGCACATGTATTTGATTGTCAAAGGTATATTCTTAAATTTGG
GTATTATTGAAAATATTTTCCATGCCTTGGTGCTAGCATATAAGTTTGGAAGT
TTGCCAACATCACAATTCATCTTGAAAAGAGCTTTTTTCCCTCCTACCACATA
CACCATTCTTAGGGAGCAATGAGGTAACAGGTCTGTGTTGTNTAGATCTTTGC
TTTTTATCCCCCTATCAGTCCAGGGCATATACTAACCTGCAAACTGATTCTGA
ATCAGGAAGGTGGTAATCAATAAGTATTCTGGCTGGGAAAGACCGTGGGCCC
AATGATCAAAGTCTTCTTGGTGCTGTTCATTAATTCTTGTGCCTTTTGGCTTGT
TTTCTAGAGTTTCTGGGCTTTGGCTGCTGATACTGCCTTTCTTAGACTGTAATT
TTTATCTGCATGCCCAGTTTCTGACCTATCAACTTGGGTTTTATTGTGCACTCT
AACTGAGCTTGTCTTCATAATTTTCTGTTTATTGCCCTGGGCTTGGATATGTCT
CAAGACACTCATGTGAATCATGCCACCCCAAATCCTGGCTTATCAAGTCCCA
GACTATAAATTATGAACTCCCATTAGCTTGGTACTAACATATACTTGATGTAG
GTATTTATGGACTTGATGATCCAAGAATATTATATTCTTCAAAATGGTTAAGC
TCCATGGAGTTAGATGACTACACTTAATGCTATTAAGTTGAACTTTTGAATGT
CAACTAATTTGCAATCAATTAAAGATACATATGCCTAGAAATTTTGAAATTTC
GGTATATTTATCCAGTTAAAGGGCTAAATTATATAAGCAAACACTACTTTTTT
TAAAACGTCTGGACTCAAAAAATGCTTTGTTCCATGTTTTAAAATTTTTAAGT
AGCAGTCTCAAAGTTGCTTAGCTGTTTATTTTGCTATGTTCCTAGCTAAGAGT
TTGGTTATAGGAGTTCATCAATAACTTATTTTTTGTACAGTTCCCACATTAGAT
ACTGTTTAAAAGTTCTTTTTTAAACTCAATTTTTTTAGAAACATAAGAGAAAT
ATTTAGATACATACAAATGTTTTTATGATTAAATAATTTTATGCTTATTTTCTG
ATACGTGTTATTTAGGTAATCATGCCCTGTACATTTAGAGGTTGCTAACTGAC
AATGTTAAGAAATTTAAAAAAAAAAAAAAAAGCCTGGGCATGATGGCTCATGC
TTGTAATCTGAACATTTGGGAGGCTGAGGCAGGAAGATCGCTTGAGGTCCAG
AGTTTAAGTCCAGCCTGGAAACATAGTTAGACCTCATCTCTACAAAAATCCTC
GTGCCG

Figure 2

page 2

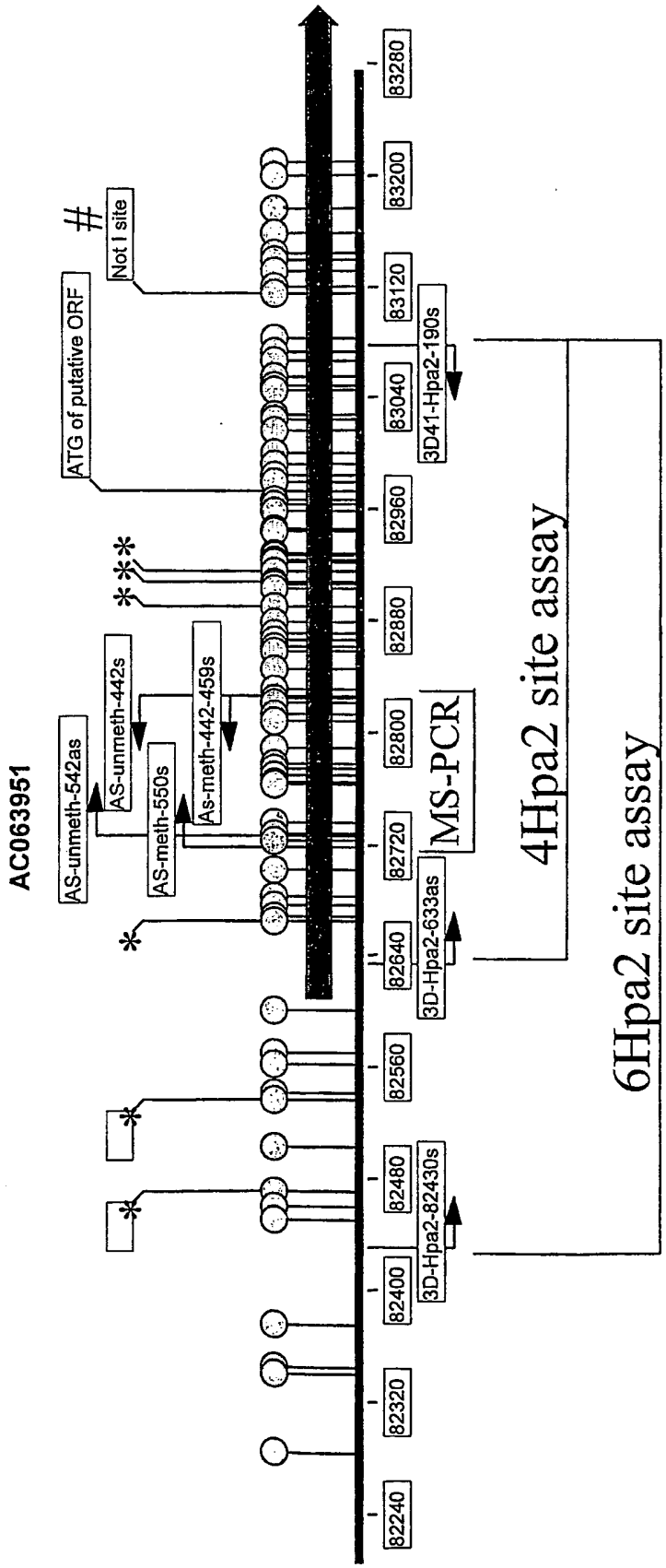


Figure 3

* Denotes a Hpa2 site

\# Denotes the Not1 site

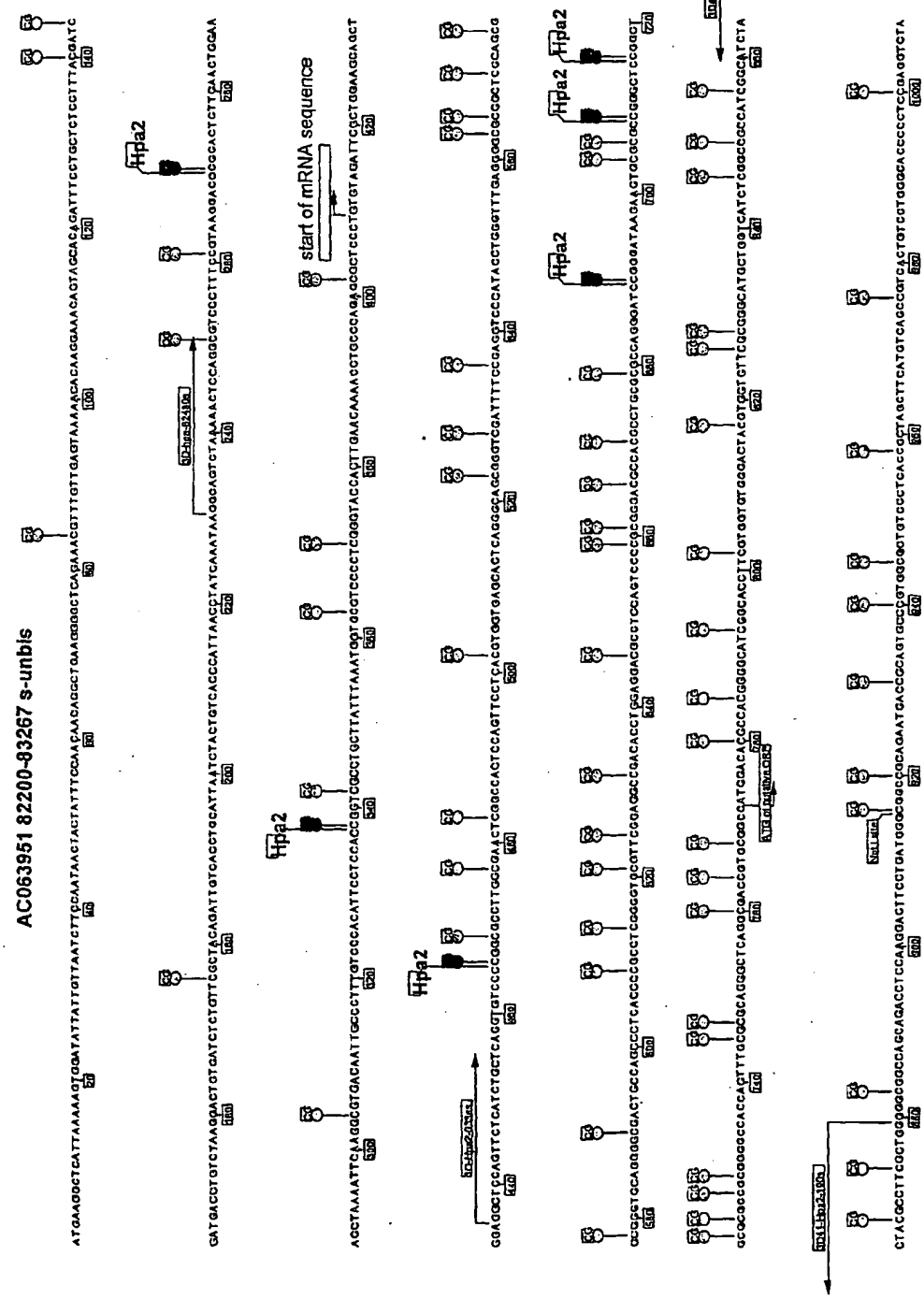

AC063951 82200-83267 s-unbis
Hpa2
start of mRNA sequence
Hpa2
Hpa2
Hpa2
Hpa2

Figure 4

Testing 4-Hpa2 and 6-Hpa2 Site Methylation Assays in Colon Cancer Cell Lines

Cell lines	4-Hpa2	6Hpa2
True positive*	22	15
False positive*	3	0
False negative	0	7
True negative*	5	8
Sensitivity	100%	68%
Specificity	63%	100%

Note: * True positive: DNA is methylated and no transcription.
False positive: DNA is methylated, but has transcription.
False negative: DNA is not methylated, but no transcription.
True negative: DNA is not methylated and has transcription.

Figure 5

Testing 4-Hpa2 and 6-Hpa2 Site Methylation Assay in Normal and Colon Cancer Tissue

Tumor/Normal status	4-Hpa2	6Hpa2
Tumor-methylated Normal-unmethylated	21	14
Tumor-methylated Normal-methylated	5	3
Tumor-unmethylated Normal-methylated	0	0
Tumor-unmethylated Normal-unmethylated	8	17
% of tumors detected	76%	50%

Figure 6

Methylation status of 62 individual CpG sites in Silenced (0ff) Colon Cancers,
Versus Non-Silenced (On) Colon Cancers and Normal Colon. Site 1 is at bp 466
And Site 61 at bp 1010 in the Wild-type Sense Sequence. Arrows indicate Hpa2
Sites at respectively bp 466, 691, 709, 716. MS-PCR1 primers amplify from sites 6-18.

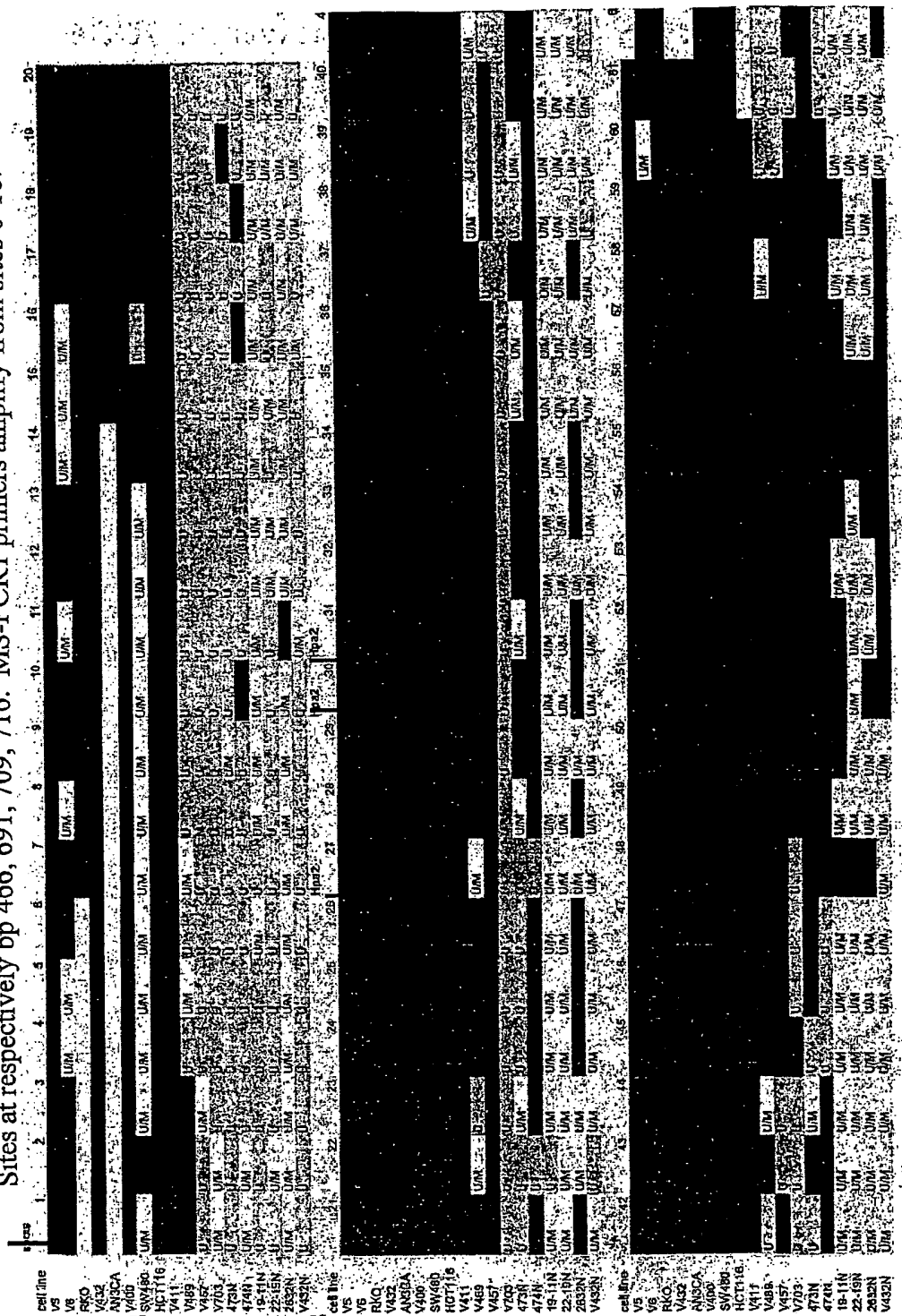
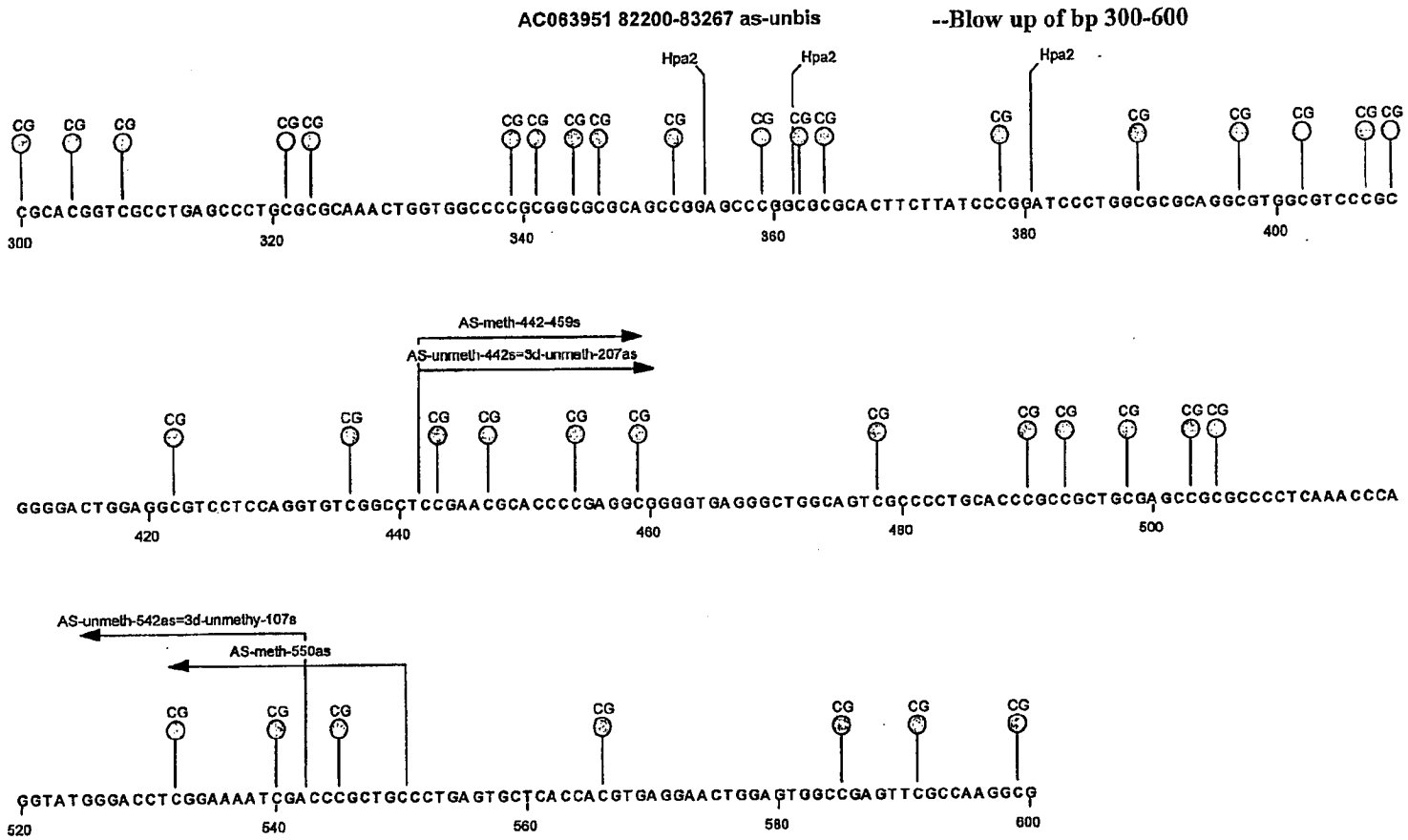


Figure 7

(Note that base pair designations run 5' to 3' on the antisense strand. Therefore The base number on the antisense strand=1068-the sense strand number.)

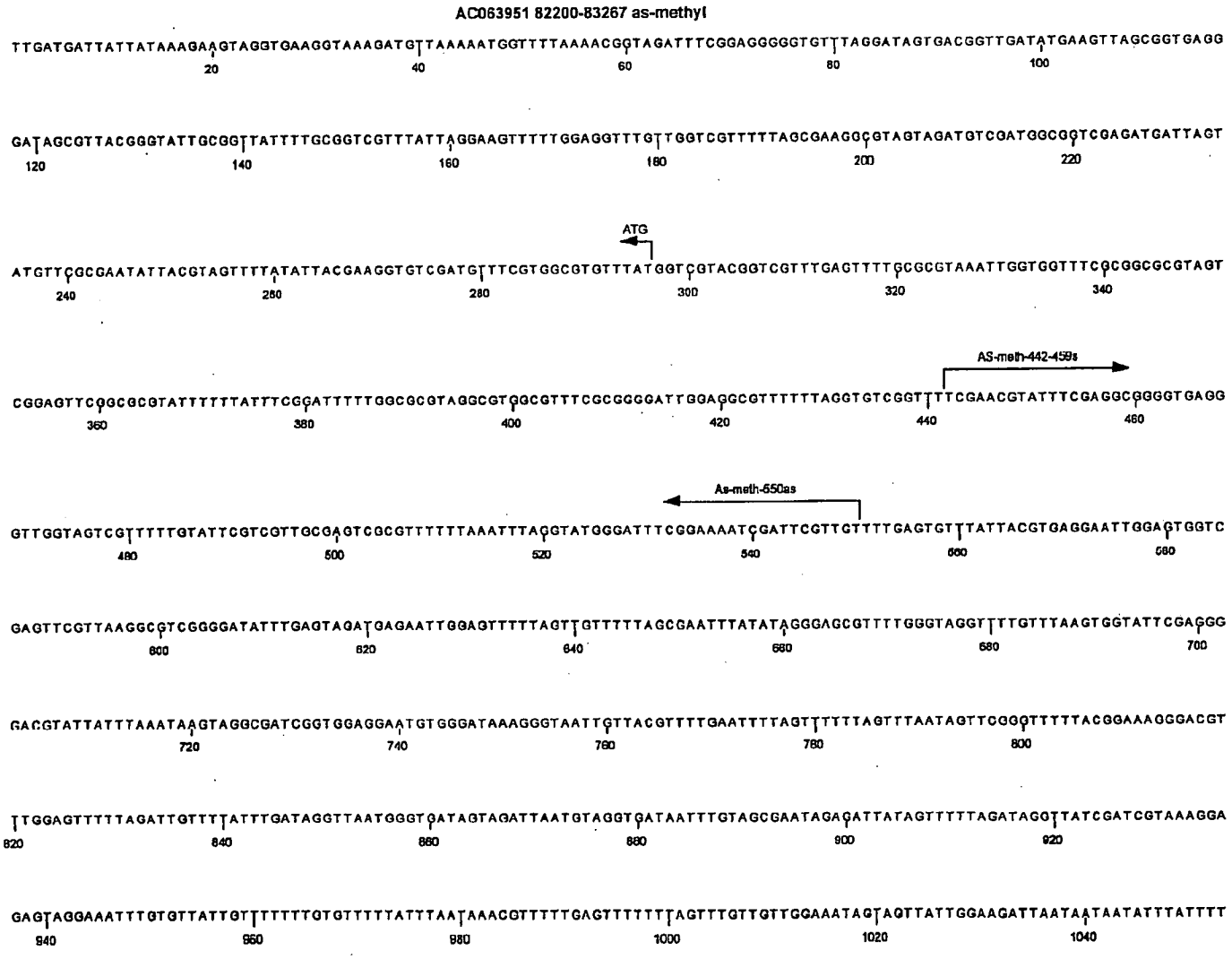
Figure 8

Antisense strand wild-type sequence showing MS and UMS PCR1 primers and CpGs



Antisense strand wild-type sequence. Arrows show positions of PCR primers specific for amplifying bisulfite converted DNA from methylated (MS-PCR1) and Unmethylated (UMS-PCR1) templates.

Figure 9

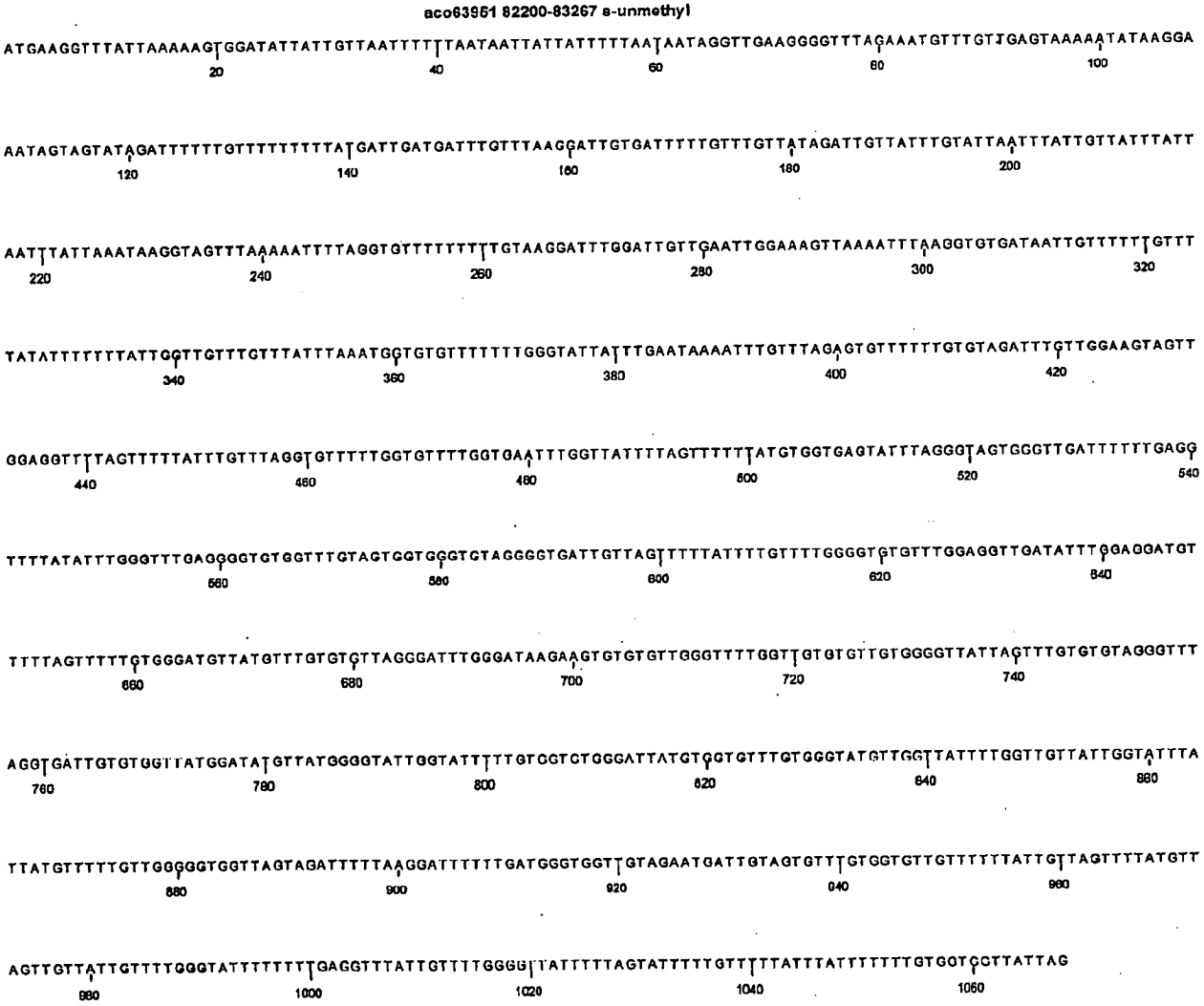


Bisulfite-converted, methyl-antisense strand, MS-PCR1□

Figure 10

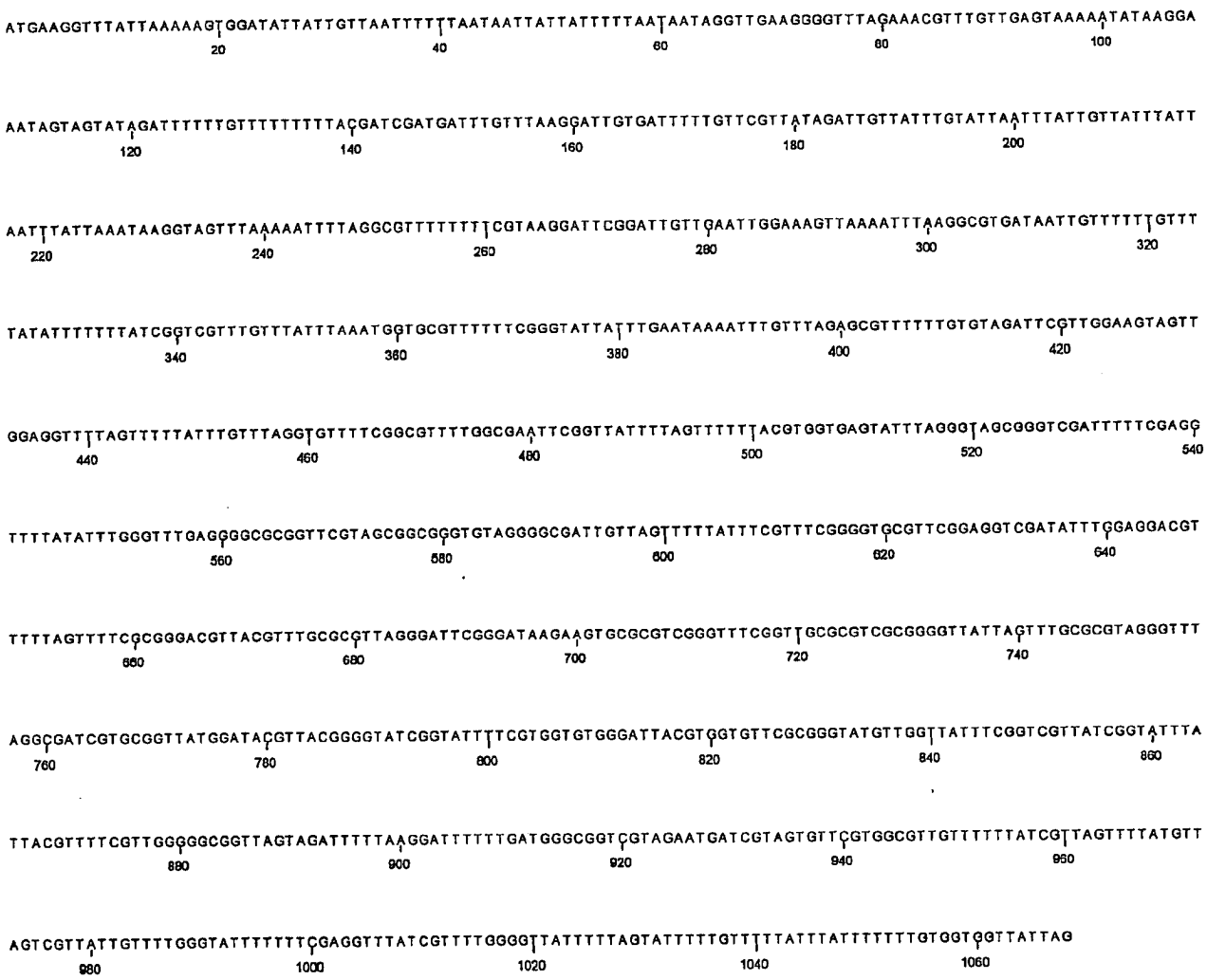
Bisulfite-converted, unmethyl-antisense strand-UMS-PCR1

Figure 11



Bisulfite converted sequence of unmethylated sense strand

Figure 12



Bisulfite converted sequence of methylated sense strand

Figure 13

Testing MS-PCR1 Methylation Assay in Colon Cancer Cell Lines

Cell lines	MS-PCR1	
True positive*	16	
False positive*	0	
False negative	7	
True negative*	8	

Sensitivity 70%
Specificity 100%

Note: * True positive: DNA is methylated and no transcription.
False positive: DNA is methylated, but has transcription.
False negative: DNA is not methylated, but no transcription.
True negative: DNA is not methylated and has transcription.

Figure 14

Testing MS-PCR1 Methylation Assay in Normal and Colon Cancer Tissue

Tumor/Normal status		
Tumor-methylated Normal-unmethylated	**34**	
Tumor-methylated Normal-methylated	**3**	
Tumor-unmethylated Normal-methylated	**0**	
Tumor-unmethylated Normal-unmethylated	**26**	
% of tumors detected	59%	

Figure 15

Testing MS-PCR1 Methylation Assay in Normal Colon Tissues in Non-Cancer Patients

Normal status		
Normal-unmethylated	**12**	
Normal-methylated	**0**	

Figure 16

Testing MS-PCR1 Methylation Assay in Colon Adenoma Tissues

Adenoma-unmethylated	**9**	
Adenoma-methylated	**19**	

% of adenomas detected 68%

Figure 17

MDTPRGIGTFVVWDYVVFAGMLVISAAIGIYYAFAGGGQQTSKDFLMGGRRMT
AVPVALSLTASFMSAVTVLGTPSEVYRFGAIFSIFAFTYFFVVVISAEVFLPVFYKL
GITSTYEYLELRFNKCVRLCGTVLFIVQTILYTGIVIYAPALALNQVTGFDLWGAV
VATGVVCTFYCTLGGLKAVIWTDVFQIGIMVAGFASVIIQAVVMQGGISTILNDA
YDGGRLNFWNFNPNPLQRHTFWTIIIGGTFTWTSIYGVNQSQVQRYISCKSRFQA
KLSLYINLVGLWAILTCSVFCGLALYSRYHDCDPWTAKKVSAPDQLMPYLVLDI
LQDYPGLPGLFVACAYSGTLSTVSSSINALAAVTVEDLIKPYFRSLSERSLSWISQ
GMSVVYGALCIGMAALASLMGALLQAALSVFGMVGGPLMGLFALGILVPFANSI
GALVGLMAGFAISLWVGIGAQIYPPLPERTLPLHLDIQGCNSTYNETNLMTTTEM
PFTTSVFQIYNVQRTPLMDNWYSLSYLYFSTVGTLVTLLVGILVSLSTGGRKQNL
DPRYILTKEDFLSNFDIFKKKKHVLSYKSHPVEDGGXDNPAFNHIELNSDQSGKS
NGTRL

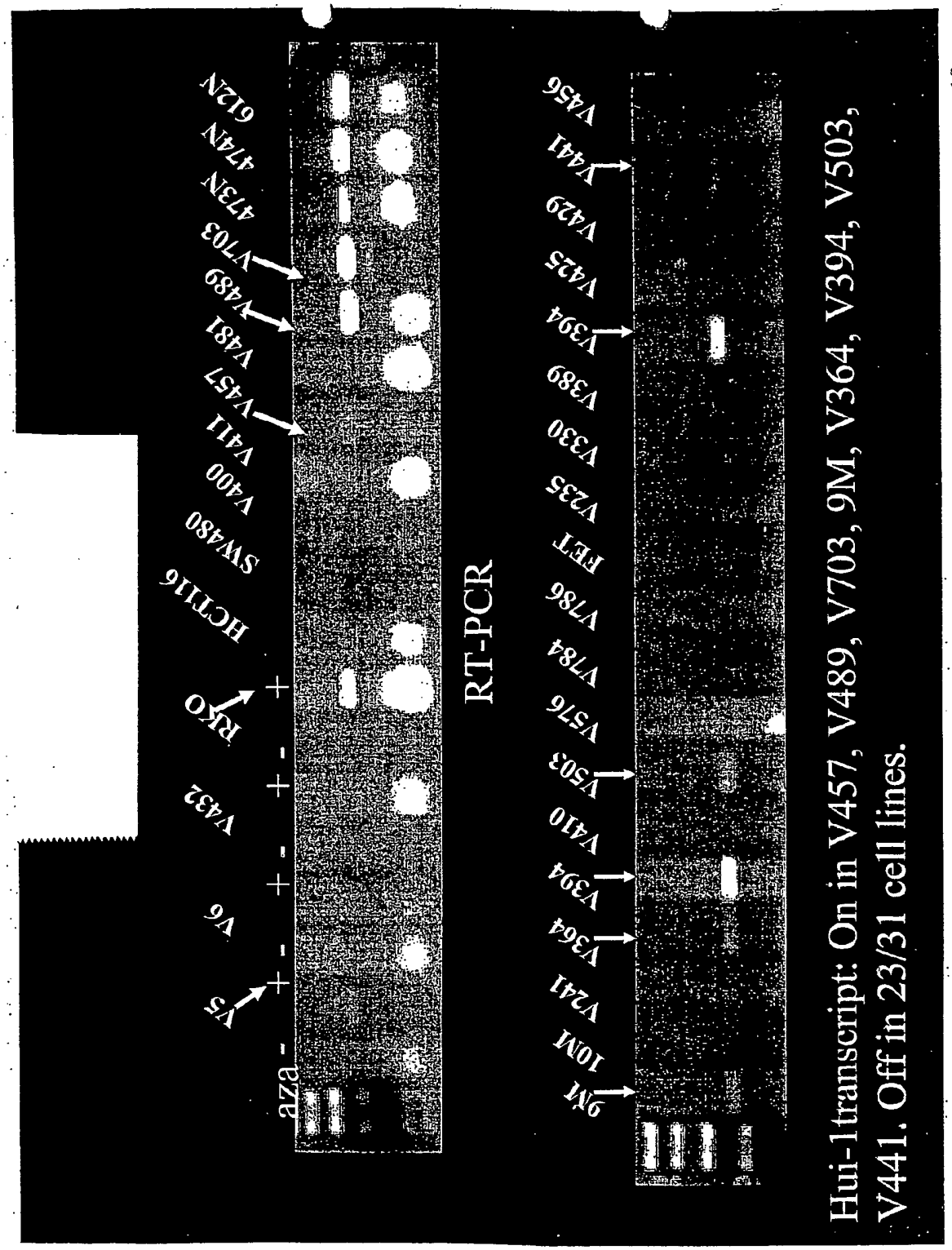

Hui-1transcript: On in V457, V489, V703, 9M, V364, V394, V503, V441. Off in 23/31 cell lines.
RT-PCR
aza - + - + - + -
V5
V6
V432
RKO
HCT116
SW480
V400
V411
V457
V481
V489
V703
473N
474N
612N
9M
10M
V241
V364
V394
V410
V503
V576
V784
V786
FET
V235
V330
V389
V394
V425
V429
V441
V456

Figure 19

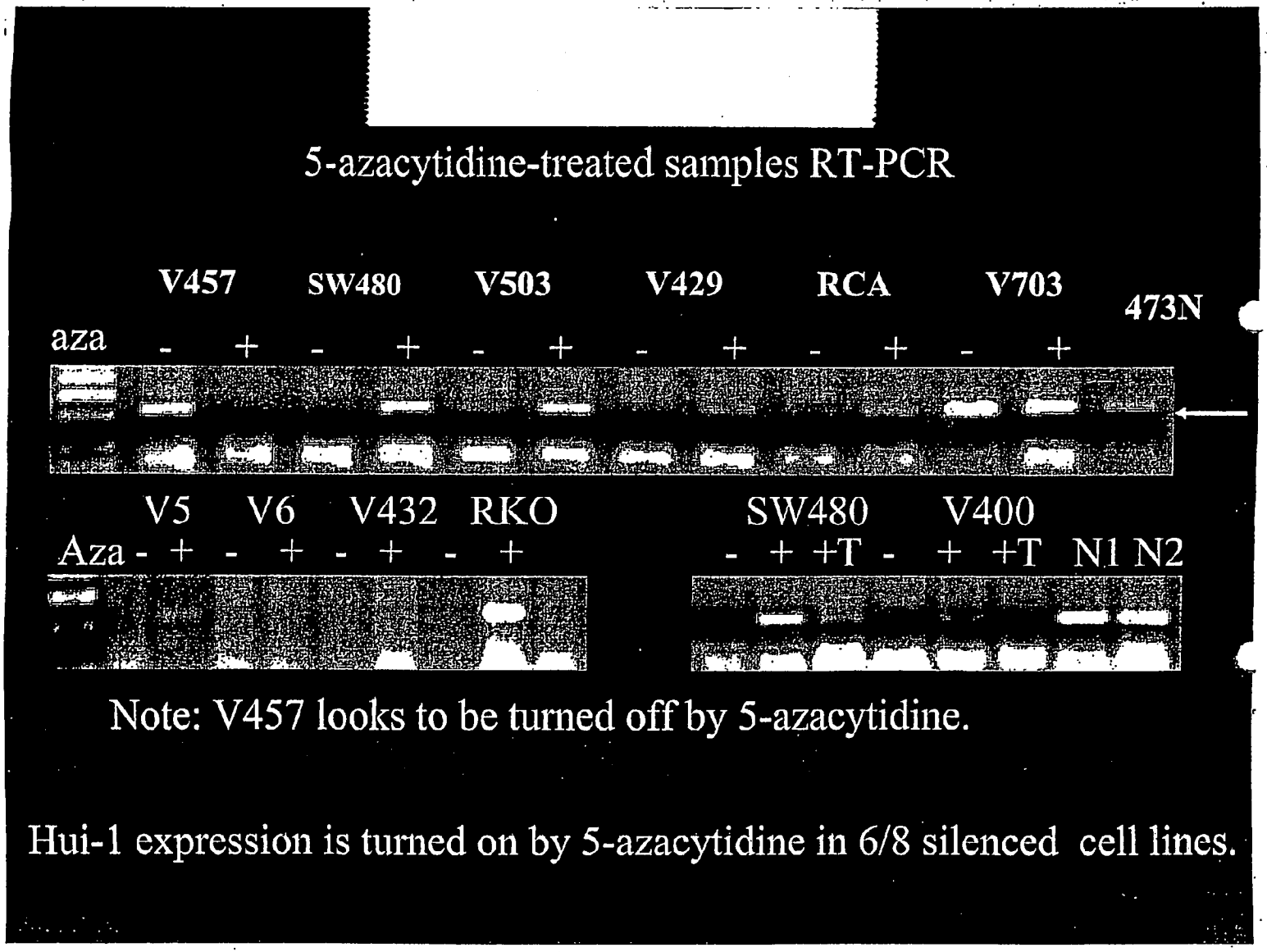
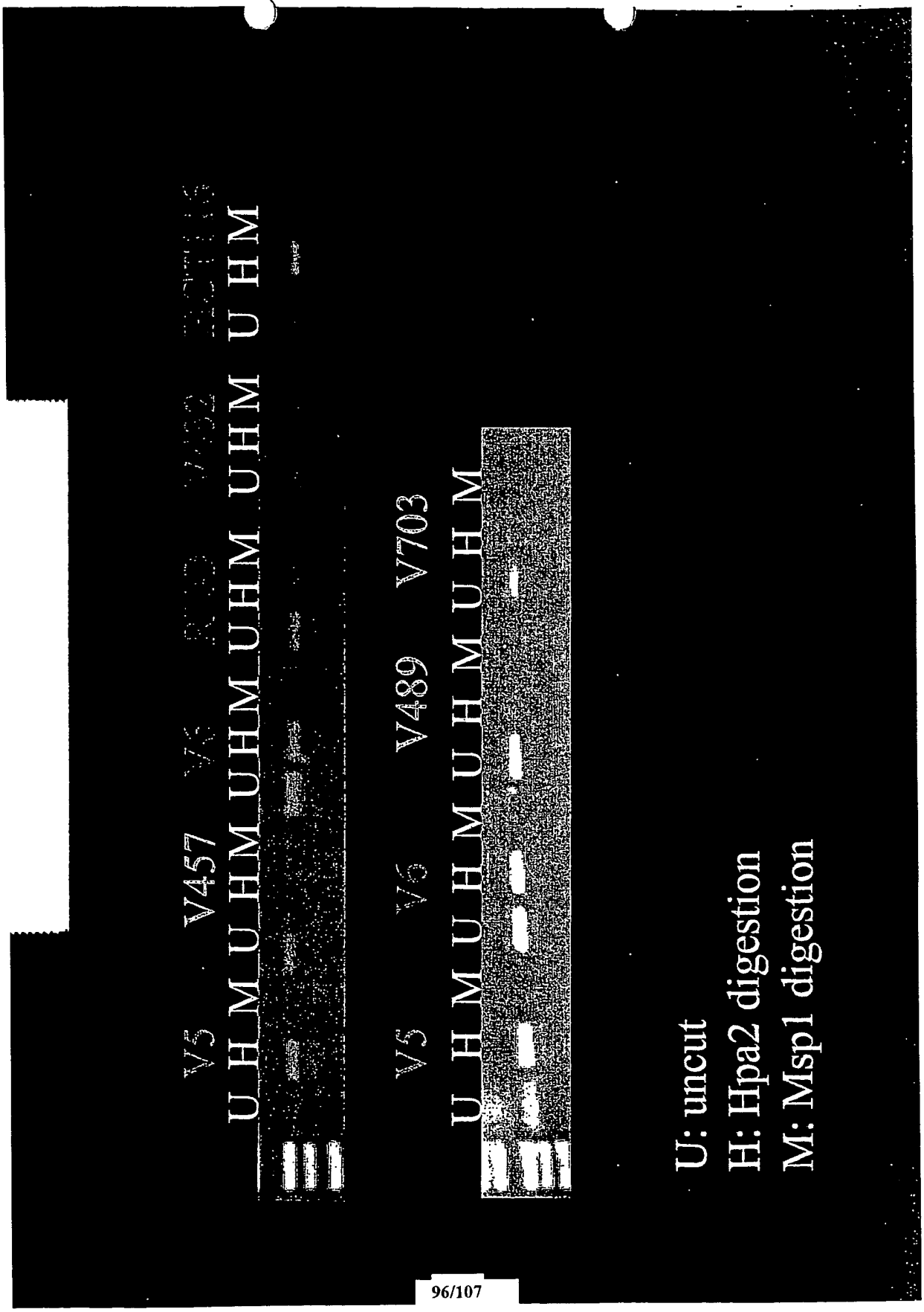

5-azacytidine-treated samples RT-PCR
V457
SW480
V503
V429
RCA
V703
473N
aza - + - + - + - + - + - +
V5
V6
V432
RKO
SW480
V400
Aza - + - + - + - + - + +T - + +T N1 N2
Note: V457 looks to be turned off by 5-azacytidine.
Hui-1 expression is turned on by 5-azacytidine in 6/8 silenced cell lines.

Figure 20


V5 V457 V5 V489 V703
U H M U H M U H M U H M U H M
V5 V5 V489 V703
U H M U H M U H M U H M
U: uncut
H: Hpa2 digestion
M: Msp1 digestion

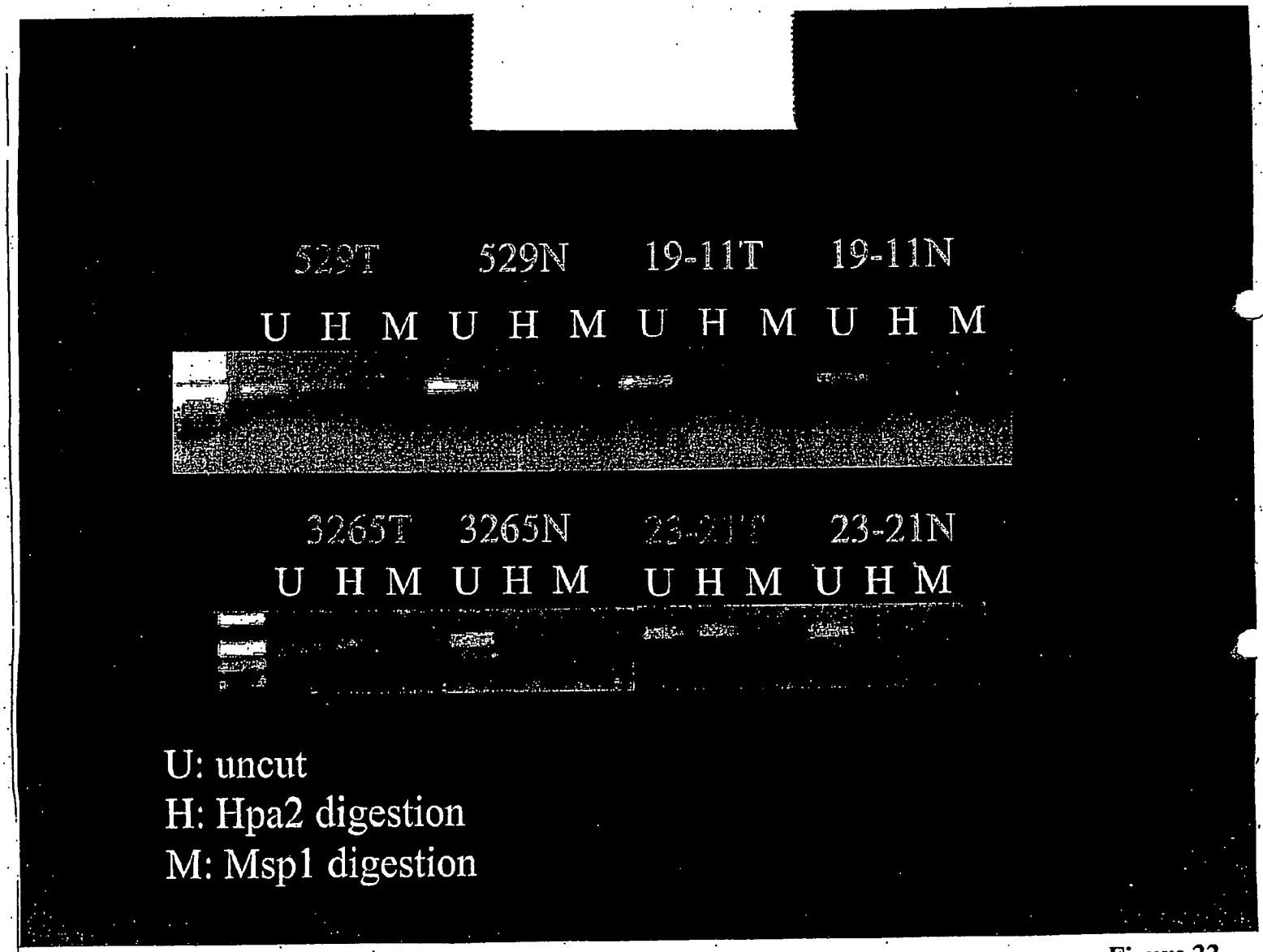

529T 529N 19-11T 19-11N
U H M U H M U H M U H M
3265T 3265N 23-21T 23-21N
U H M U H M U H M U H M
U: uncut
H: Hpa2 digestion
M: Msp1 digestion

Figure 22

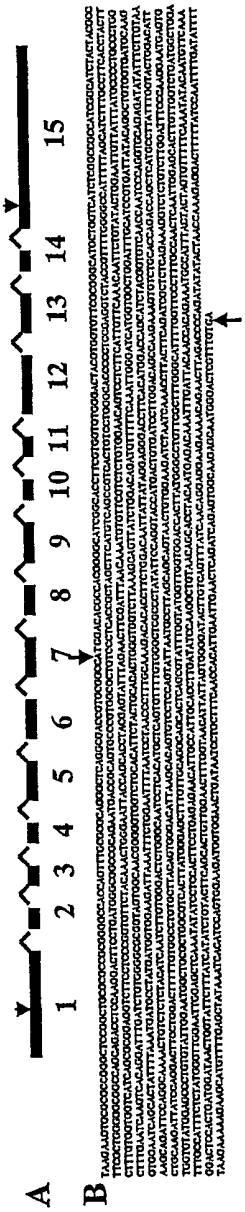


Figure 23

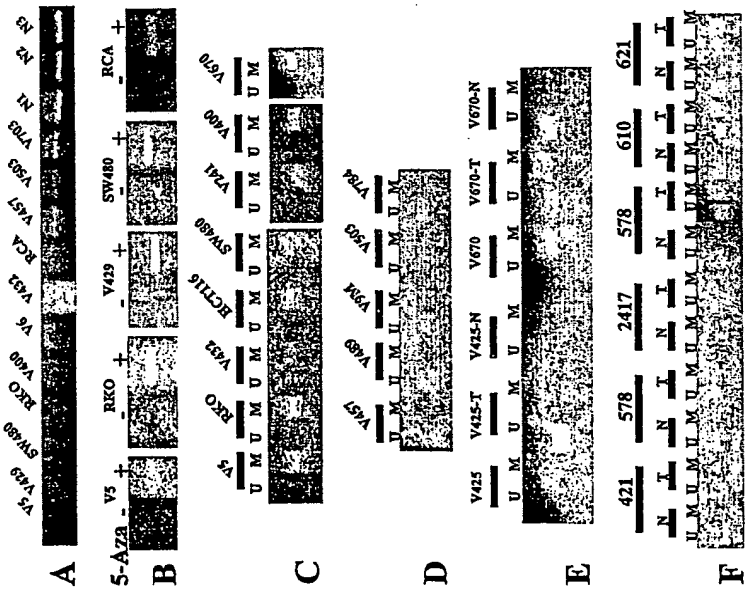

A
V5
V425
SW480
RKO
V400
V6
V423
RCA
V427
V503
V703
N1
N2
N3
B
5-Aza
V5
RKO
V429
SW480
RCA
C
U M U M U M U M U M U M U M U M U M U M
V5
RKO
V423
HCT116
SW480
V641
V400
V670
D
V427
V689
V9X
V503
V704
E
V425
V425-T
V425-N
V670
V670-T
V670-N
F
421
578
2417
578
610
621
N T N T N N T N T N T
U M U M U M U M U M U M U M U M U M U M U M U M

Figure 24

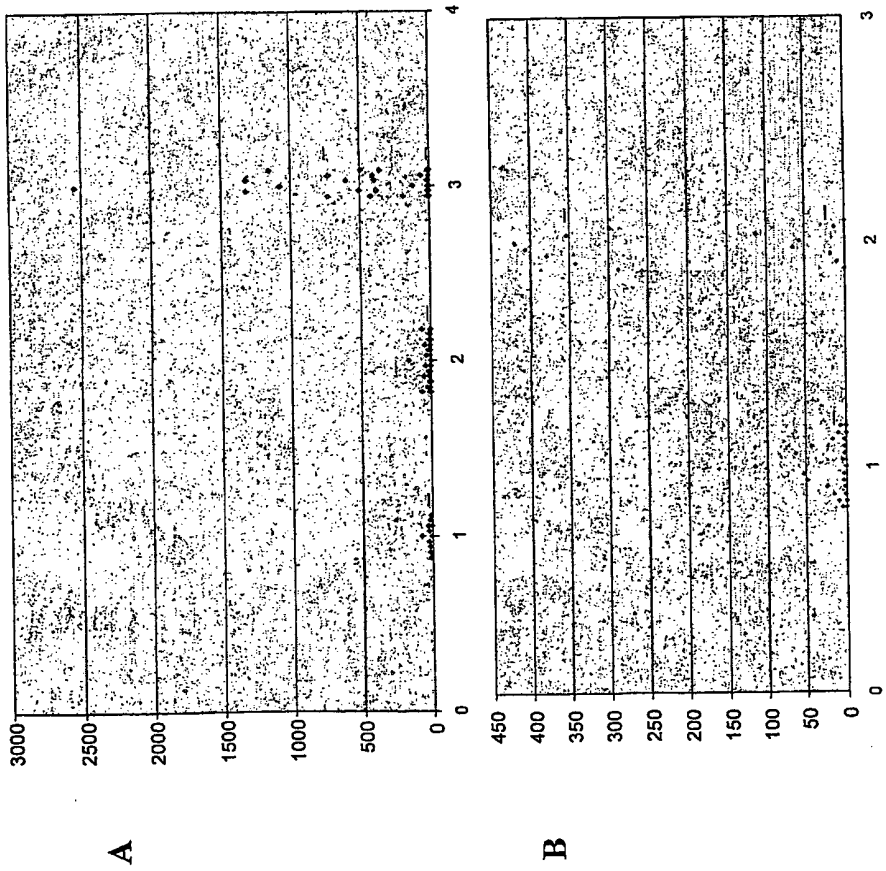

(SLC5A8/MYOD1) *1000
A
3000
2500
2000
1500
1000
500
0
0
1
2
3
4
B
450
400
350
300
250
200
150
100
50
0
0
1
2
3

Figure 25

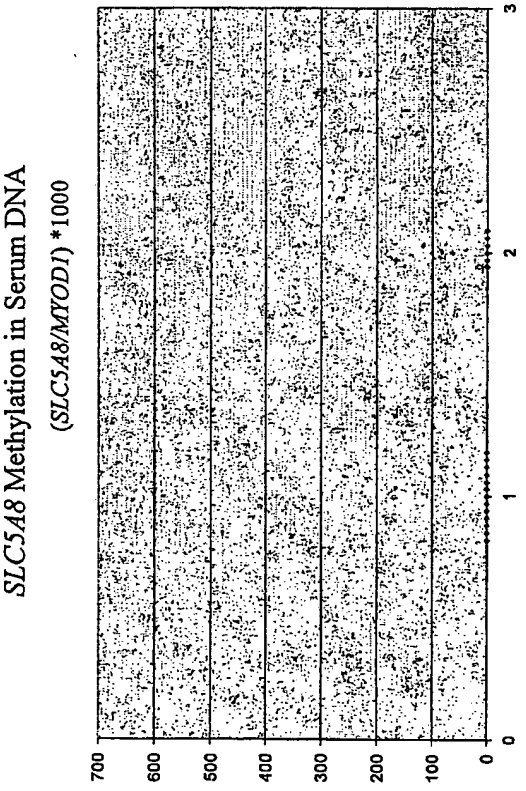

SLC5A8 Methylation in Serum DNA
(SLC5A8/MYOD1) *1000
700
600
500
400
300
200
100
0
0
1
2
3

Figure 26

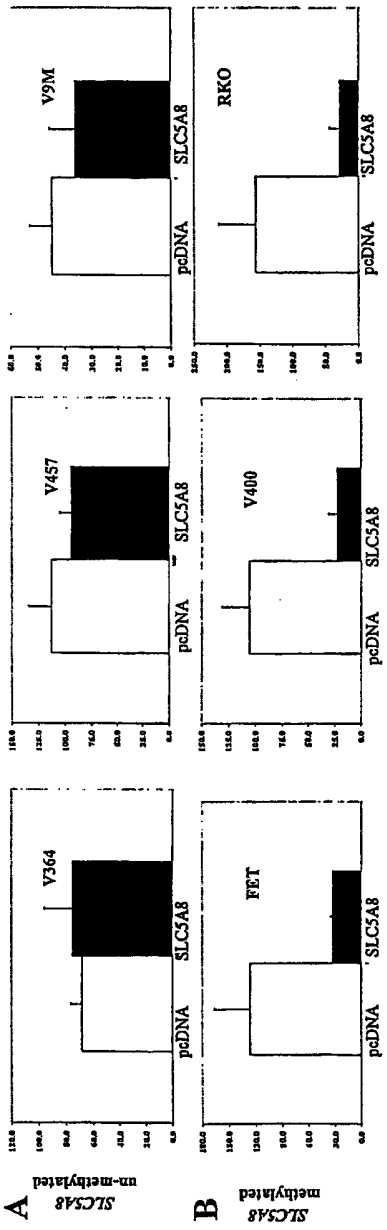

A
un-methylated
SLC5A8
V364
pcDNA
SLC5A8
V457
pcDNA
SLC5A8
V9M
pcDNA
SLC5A8
B
methylated
SLC5A8
FET
pcDNA
SLC5A8
V400
pcDNA
SLC5A8
RKO
pcDNA
SLC5A8

Figure 27

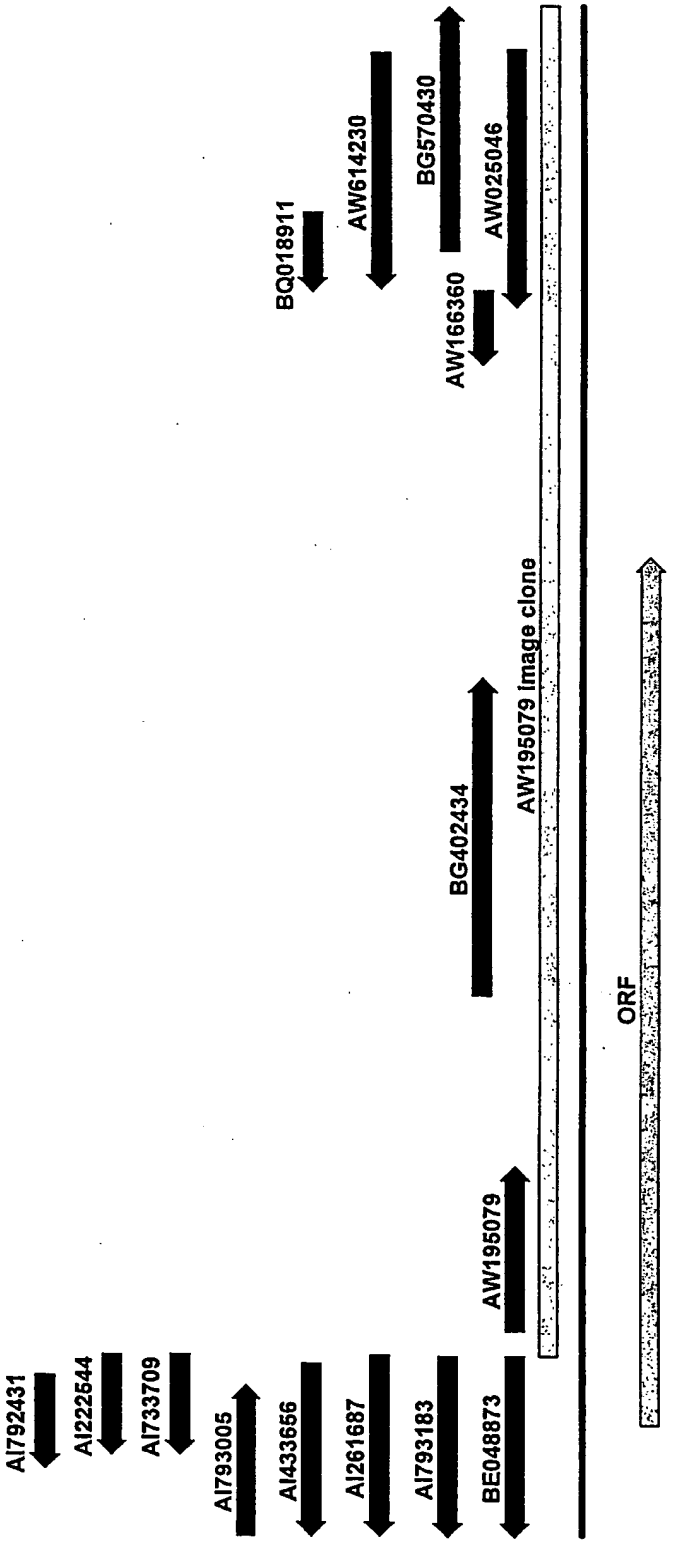

AI792431
AI222544
AI733709
AI793005
AI433656
AI261687
AI793183
BE048873
AW195079
BG402434
AW195079 image clone
ORF
BQ018911
AW614230
BG570430
AW166360
AW025046

Figure 28

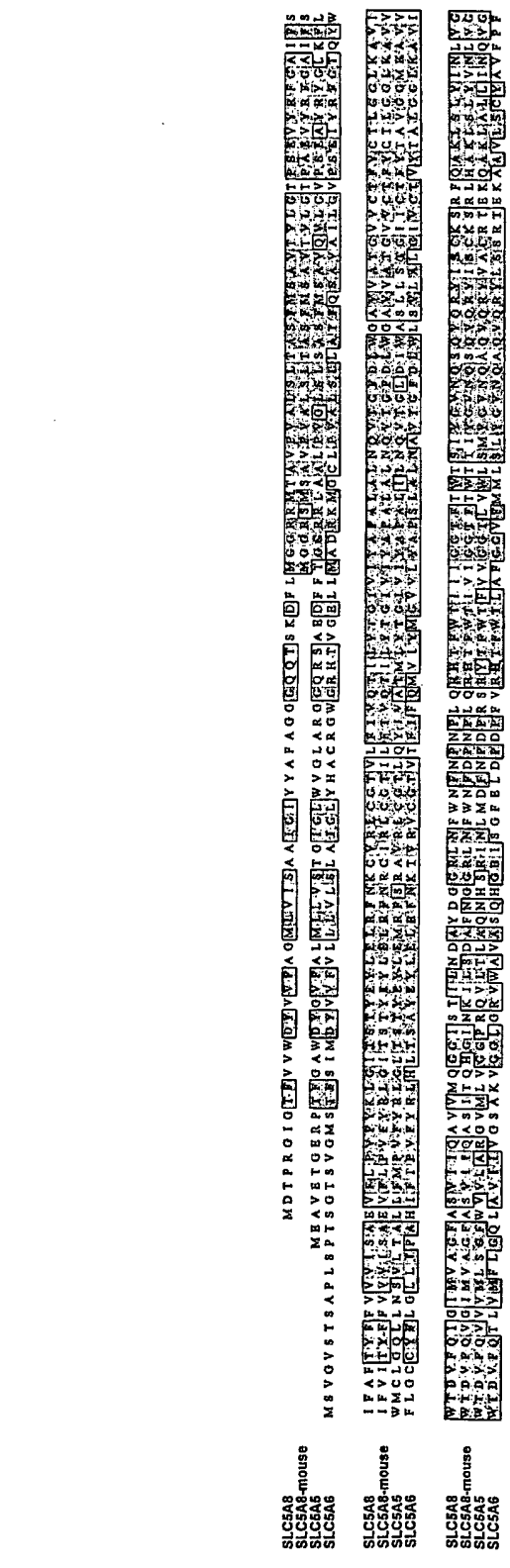


Figure 29


A
ATG
Not1
82680
82720
82760
82800
82840
82880
82920
82960
83000
83040
83080
83120
83160
83200
B
SAMPLES
V5
V432
V400
HCT116
V411
V6
SW480
RKO
AN3CA
N-V432
473N
1911N
2219N
474N
2632N
V457
V703
V489
off
on

Figure 30

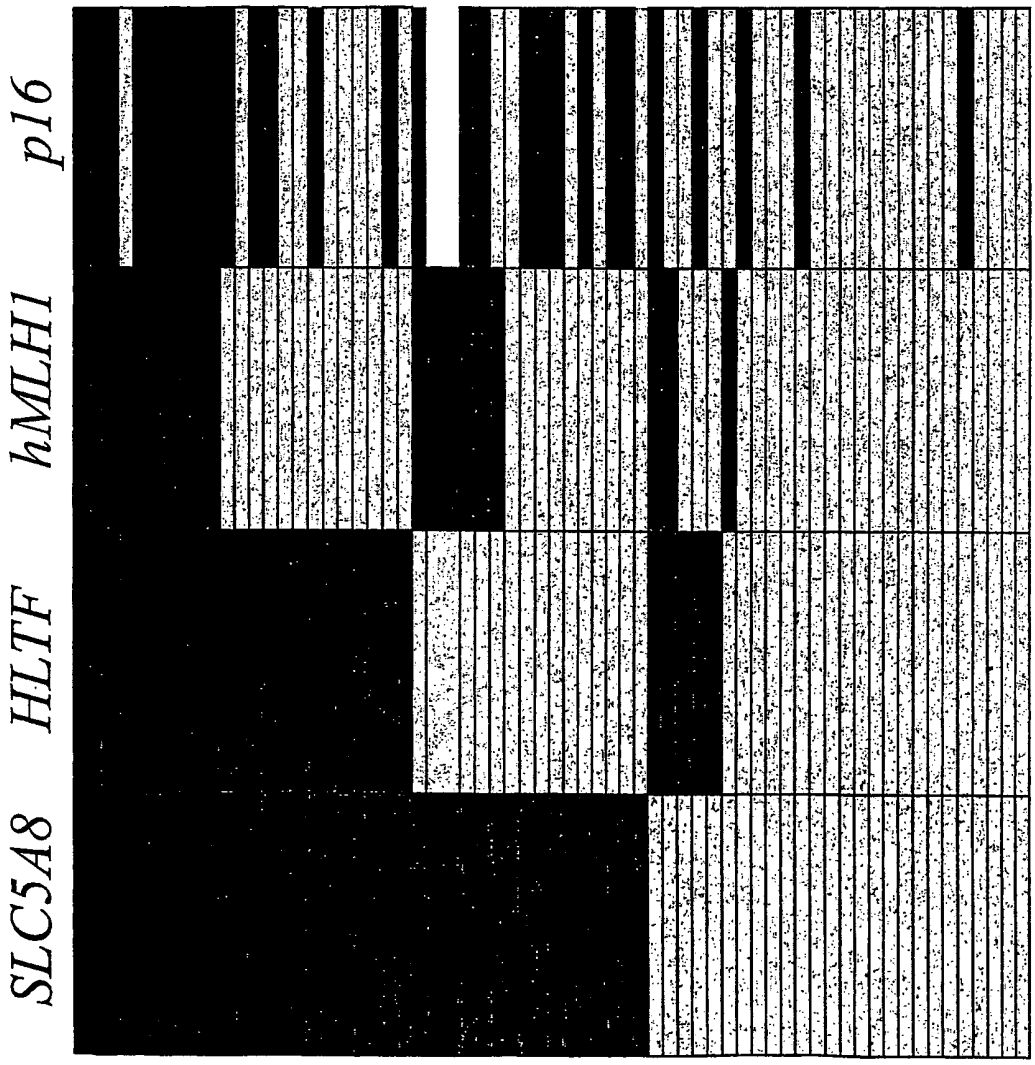


Figure 31

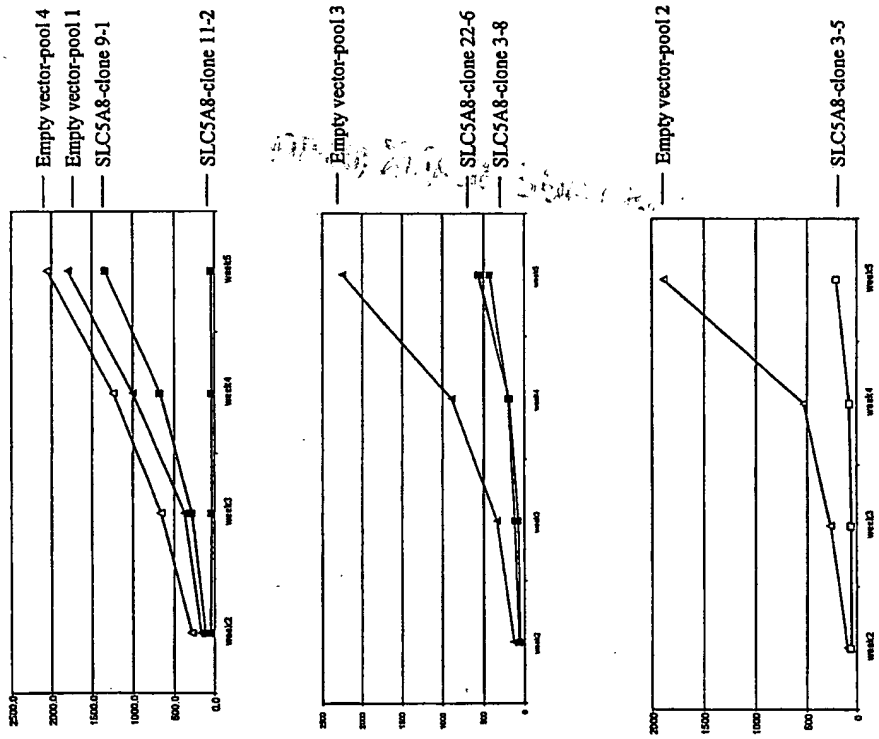

Empty vector-pool 4
Empty vector-pool 1
SLC5A8-clone 9-1
SLC5A8-clone 11-2
Empty vector-pool 3
SLC5A8-clone 22-6
SLC5A8-clone 3-8
Empty vector-pool 2
SLC5A8-clone 3-5

Figure 32